QUANTITATIVE ANALYSES OF TRAIN DERAILMENT PROBABILITY AT HIGHWAY-RAIL GRADE CROSSINGS

BY

SAMANTHA CHADWICK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Christopher P.L. Barkan, Chair
Dr. M. Rapik Saat, Association of American Railroads
Professor Rahim Benekohal
Professor Paolo Gardoni
Professor Ian Savage, Northwestern University

ABSTRACT

The current methodology for prioritizing highway-rail grade crossing (HRGC) warning system upgrades focuses on the likelihood of collisions and highway user casualties at crossings. However, these two metrics do not encompass all crossing risks. Specifically, they do not consider the potential for grade crossing incidents to cause train derailments and the consequent casualties to passengers and crew members, property damage, and release of hazardous materials. In contrast to the large body of research devoted to understanding the impact of crossings on highway users, almost no research has considered the risk that highway users pose to trains at HRGCs. With increased interest in passenger rail transport and the growth in transportation of hazardous materials such as crude oil, the importance of a comprehensive understanding of the risk of HRGC collisions is critically important.

This dissertation develops an HRGC-caused derailment probability calculator using data analytics and statistical modeling. The Federal Railroad Administration (FRA) and state Departments of Transportation (DOTs) have developed large databases of historical incidents that can be used to better understand the effect HRGCs have on train derailment rates. I use these databases to develop statistical regression models that quantify actual experience to understand the differences between derailment and non-derailment incidents. I first develop a set of univariate statistical analyses to identify the incident-specific factors affecting derailment likelihood. Then, I develop three logistic regression models of derailment likelihood with these factors as input variables. Next, I develop a series of proxy variables to relate the incident-specific factors to crossing-specific characteristics. All of this is combined in a spreadsheet-based calculator, whose function I illustrate with a case study of four Illinois rail corridors. I

combine these results with incident likelihood predictions generated by the FRA's WBAPS system to show how consideration of derailment likelihood can affect crossing prioritization.

By quantifying derailment likelihood, my research adds a new dimension to our understanding of how to assess grade crossing risk and warning system upgrade prioritization. The model allows users to identify crossings with high derailment likelihood, something that was not previously possible. This model will enable more informed allocation of safety resources to minimize the occurrence of derailments at grade crossings. It can be integrated into an overarching risk analysis framework that would consider all sources of risk at a grade crossing. Ultimately, this tool will open up new opportunities for railroad risk reduction, leading to a safer operating environment for railroads, rail passengers, highway users, and the general public alike.

ACKNOWLEDGEMENTS

I met Chris Barkan nearly a decade ago, completely by chance, when I was a sophomore in civil engineering at UIUC. At that time, had you asked me if I planned on pursuing a Ph.D. in railroad engineering, the answer would have been a confused “what?”, yet somehow, here we are.

This has been a long and at times very difficult process, which I only survived thanks to the help of numerous people along the way. First and foremost, I would like to thank Chris for his ongoing advice, encouragement, and teaching, as he has put me firmly on the path to a fulfilling career in a field I love. I am very fortunate to have had the opportunity to work with him over the years, as his group attracts a variety of people who are not only excellent engineers, but also kind and generous individuals. I have benefitted from the many talented RailTEC faculty and research engineers who have advised me and guided my research, especially Rapik Saat and Tyler Dick who provided critical input in the development of this research. I also want to acknowledge the amazing RailTEC staff, including LB and Emma, without whose assistance in many matters it would have been impossible to get to this point in my academic career.

Funding for this research was provided by the Association of American Railroads, the ABS Group, and the National University Rail Center, a US DOT OST Tier 1 University Transportation Center. I was also supported by a CN Graduate Research Fellowship in Railroad Engineering and a U.S. DOT Eisenhower Transportation Fellowship. I am grateful for their generous financial support and their recognition of my work. I would also like to acknowledge Robert Fronczak and Mike Martino at the AAR, who helped obtain support for this research and provided insight and feedback over the years.

Through conferences and presentations, I have met many experts in the rail safety field whose feedback has greatly improved my work. I would especially like to thank Ian Savage from Northwestern University, who served on my doctoral committee, along with Rapik Saat, Rahim Benekohal and Paolo Gardoni. I am grateful to Steve Laffey from the Illinois Commerce Commission who showed great interest in my research and provided assistance and valuable insight on the Grade Crossing Inventory database, as well as detailed Illinois crossing data. They, and other colleagues I have met over the years, contribute to the cooperative, positive environment that characterizes the field of rail safety research.

On a personal level, it genuinely took a village to get me to this point, and if I were to name all the friends and colleagues who supported me over the years, this dissertation would be another 300+ pages. I would like to thank all my fellow RailTEC students over the years; it has been a tremendous pleasure getting to know and work with so many great people, and I am excited to continue working with them as our careers develop. I'd also like to thank my close undergrad friends, my gaming group and my trivia team, friends who have all (sometimes literally) provided a shoulder to cry on.

I owe most of who I am today to my amazing family. My brother Alex is a constant inspiration for me, especially as our grad school careers have overlapped for the past four years and I have seen how hard he works on his own research. (I'm sorry I will beat you to being Dr. Chadwick.) We inherited our smarts from our parents, from whom I still learn something new every day. I am grateful to them for instilling in me a love of knowledge, a healthy willingness to take risks, and a taste for adventure. Thank you both so much for supporting me in all my ventures, and especially in this one – you have been an endless source of ideas and encouragement. And thank you, Mom, for turning your expert eye to this whole document!

Last but far from least, I'd like to thank my wonderful fiancé Mike, who has been my rock for the past seven years while working on his own Ph.D. at Purdue. I'm not sure how we managed to complete two doctorates and still fit in so much fun and adventure. I'm glad I got to go through this journey with you.

To my family, especially:

Grandpa Glenn (a civil engineer), Gramps (a Dr. Chadwick)

Grandma, and Grammy (two smart and strong women)

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1. BACKGROUND

Highway-rail grade crossing safety has been a topic of concern to both railroads and the general public since the earliest days of railroads. For most of the first century of railroading, grade crossings were for horse-drawn vehicles. In the first decades of the 20^{th} century, incidents with motor vehicles increased. With the rise of the personal automobile, by the 1920s over 1,000 people were being killed at crossings each year, with an additional 4,000 injuries (Aldrich, 2008; Enoch, 2014). Today, with about 2,000 incidents per year, grade crossing collisions are the most common cause of railroad incidents by a substantial margin. They are also the leading cause of railroad-related casualties (about 1,000 per year) and the second-largest cause of railroad-related fatalities (about 250 per year) (FRA, 2011a).

From 1991 to 2010, approximately 71,000 collisions occurred at public highway-rail grade crossings in the United States, including about 57,000 at publicly-accessible grade crossings on mainline railroad tracks (FRA, 2011a). The next most frequent incident cause is broken rails/welds, which led to approximately 2,200 incidents over the same 20 years.

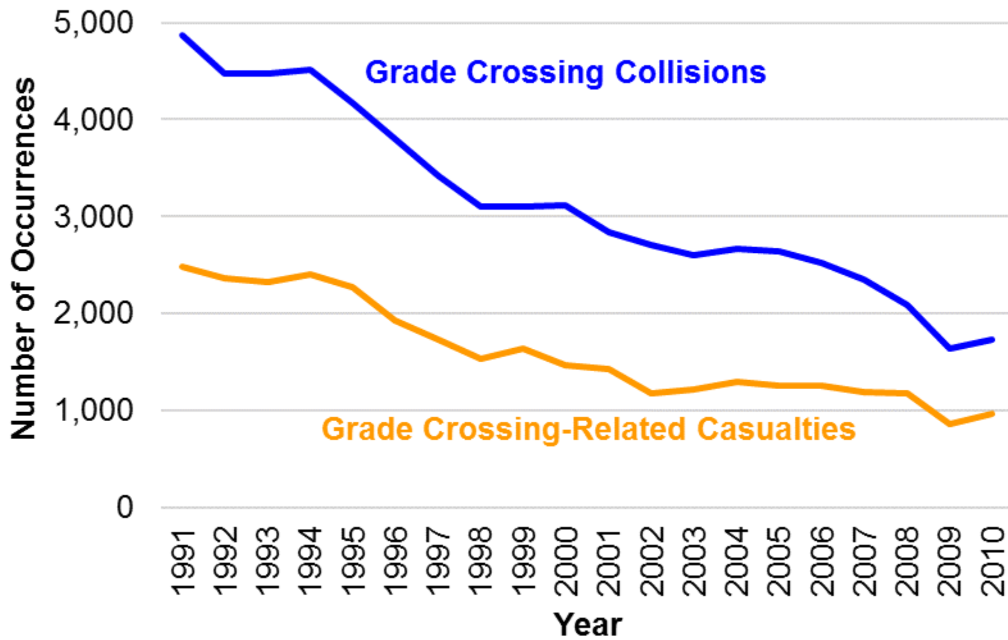
Much of the attention to improving grade crossing safety has been motivated by concern for highway user safety. However, these collisions also have the potential to cause train passenger and crew casualties, property damage, and the release of hazardous materials. There have been several serious grade crossing collisions in recent years. In 2015 alone, passenger trains were involved in several casualty-causing grade crossing incidents including in Valhalla, New York; Oxnard, California; and Halifax, North Carolina (NTSB, 2015a, 2015b; Associated Press, 2015). Recent grade-crossing-caused derailments involving freight trains have also resulted in casualties and major property damage. An incident in Rosedale, MD in 2013 resulted in the release of four cars of hazardous materials (sodium chlorate crystal and terephthalic acid), leading to a fire and large explosion (NTSB, 2014). Such incidents, as well as substantially expanded rail transport of flammable liquids have led to renewed interest in developing new grade crossing management strategies (NCHRP, 2014).

The U.S. railroad network consists of approximately 140,000 miles of track, and in 2016 there were approximately 220,000 publicly-accessible highway-rail grade crossings (FRA, 2016). This corresponds to a national average of almost one grade crossing per half mile (AAR, 2017) and in urban centers, the density of crossings can be much greater. Furthermore, in these urban areas, highway vehicle traffic may be particularly high, and the population density of areas adjacent to the crossings and thus potentially affected by an incident is also greater. Resources for highway-rail grade crossing improvements are limited, so it is in the interest of both the private and public sector to identify and rank which crossings pose the greatest risk. Consequently, a comprehensive understanding of all factors affecting this risk is necessary so that resources can be invested most efficiently and effectively.

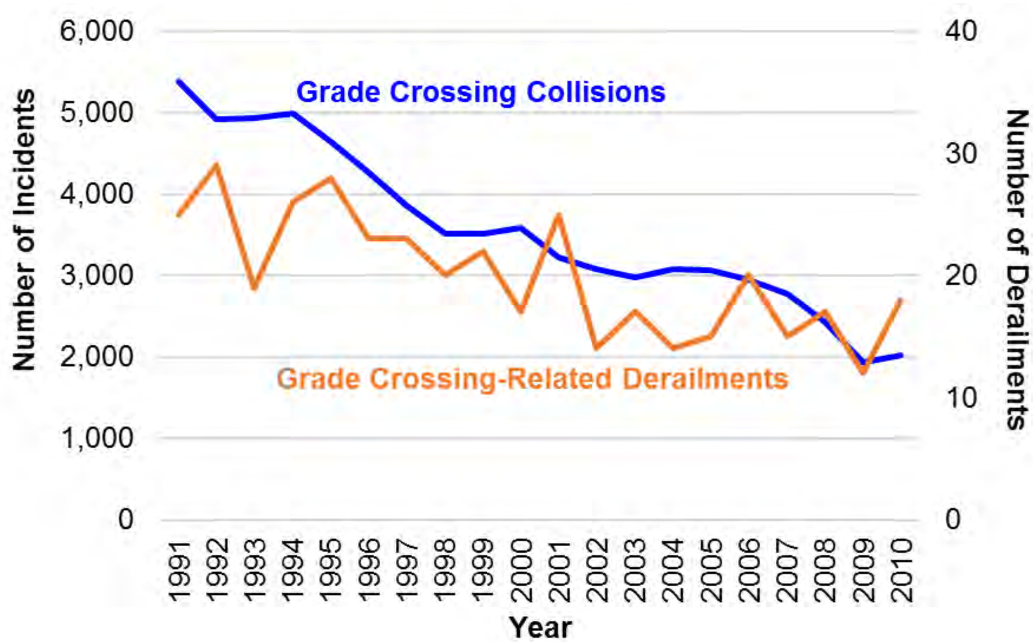
As stated above, grade crossings have long been understood to pose risk to highway users. Early campaigns to improve safety led to standardized signage at grade crossings, and expanded educational initiatives. The number of fatalities at grade crossings in the 1980s was around 600 per year (FRA, 2017a). In an effort to reduce this number, railroads, government and non-government organizations, and researchers devoted significant effort and resources to reduce the risk. This has significantly improved grade crossing safety, with fatalities declining to between 200 and 300 fatalities each of the past five years (FRA, 2017a).

A variety of methods for modeling collision likelihood have been developed, focusing on the risk trains pose to highway vehicles and their occupants, including the widely-used U.S. Department of Transportation Accident Prediction Model (FRA, 1987; Ogden and Korve Engineering, 2007). In addition a variety of models have been developed to address limitations of that model (Benekohal and Elzohairy, 2001; Austin and Carson, 2002; Saccomanno et al., 2004; Oh et al., 2006; Washington and Oh, 2006; Saccomanno et al., 2007). The results of these and other studies have led to improved grade crossing warning systems, integration of grade crossing operations with highway traffic signaling, public education programs such as Operation Lifesaver, and numerous other improvements in engineering and education (Mok and Savage, 2005). These technologies and programs aim to reduce the number of casualties due to train-highway vehicle collisions, and the result has been a steady decline in the number of incidents and casualties over the past several decades (Figure 1.1a).

Although the focus on grade crossing safety has led to considerable improvements, one aspect has been largely overlooked – the risk that highway-rail grade crossings pose to trains. While the number of grade crossing collisions each year has been going down, the number of these incidents that result in derailment has not seen a similar decline (Figure 1.1b; Table 1.1). Each year, 0.4 to 1% of grade crossing collisions result in a train derailment, with the rate appearing to rise slightly in the past five years (Figure 1.1c; Table 1.1).

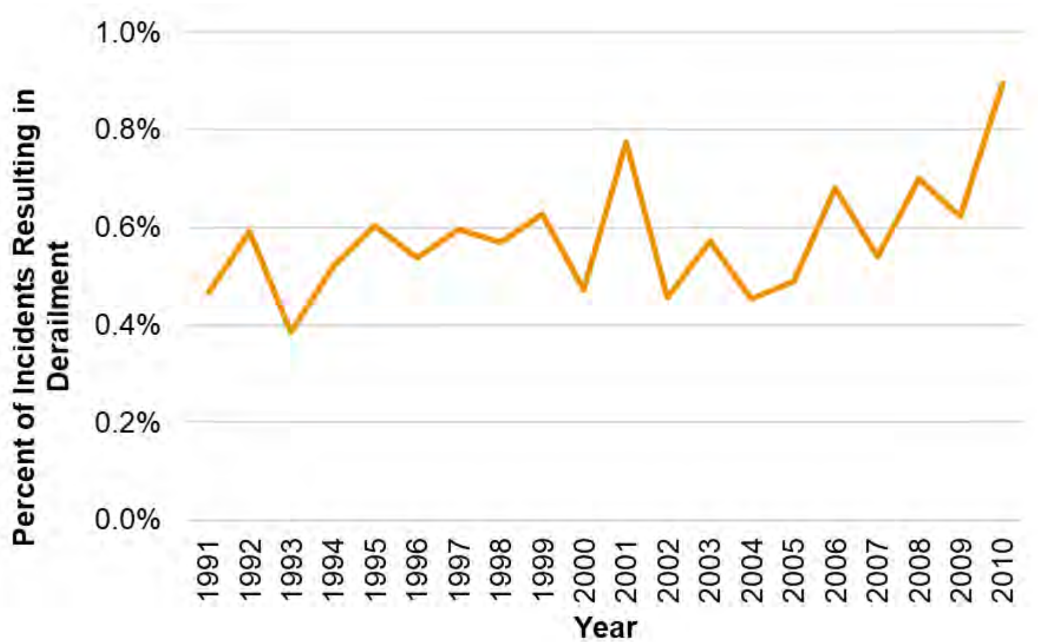


(a)



(b)

Figure 1.1: (a) Trend in number of grade crossing collisions and casualties over time (Zhang 2017). (b) Trend in grade crossing collision-caused derailments over time.



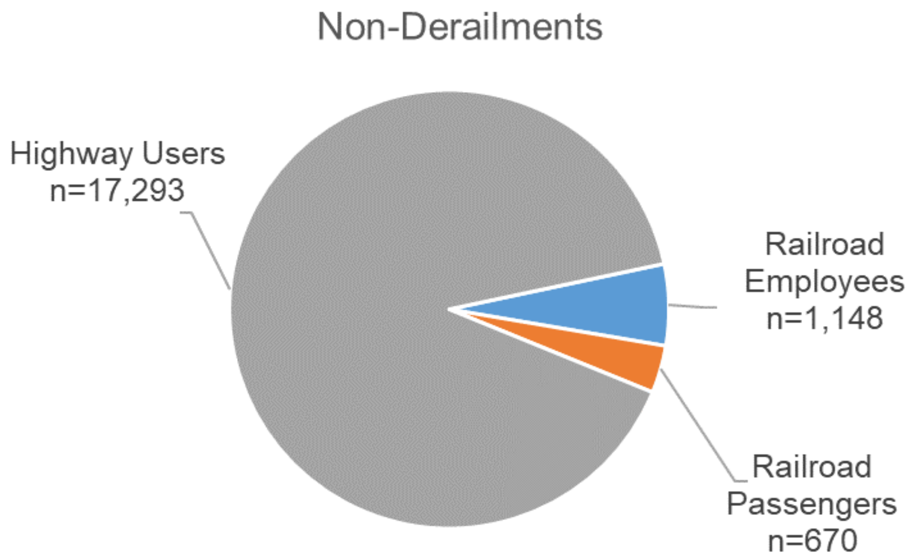
(c)

Figure 1.1 (cont): (c) Percentage of grade crossing incidents that resulted in derailment, 1991-2010.

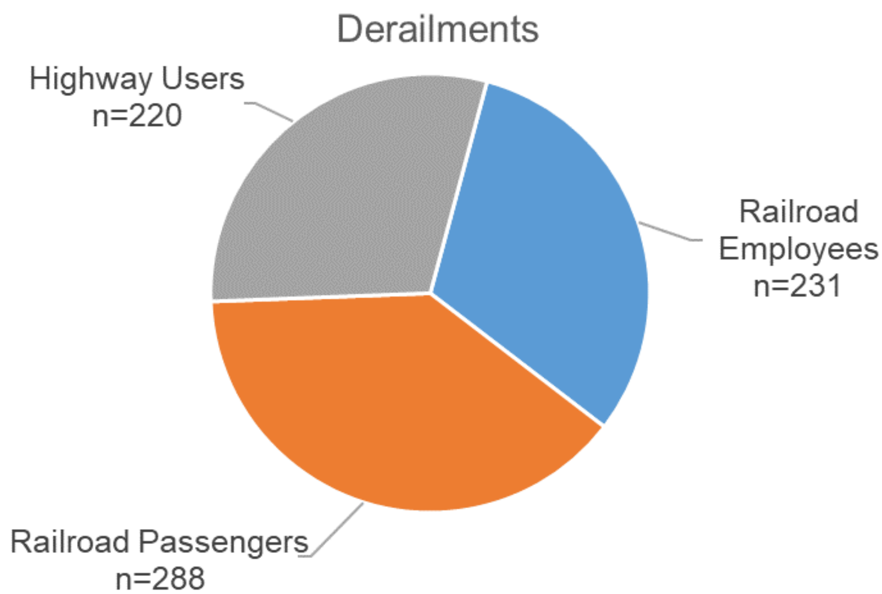
Table 1.1: Number of Grade Crossing Collisions and Related Derailments per Year, 1991-2010

Year	Grade Crossing Collisions	Grade Crossing-Related Derailments	Derailment Rate
1991	5387	25	0.46%
1992	4917	29	0.59%
1993	4933	19	0.39%
1994	4988	26	0.52%
1995	4645	28	0.60%
1996	4268	23	0.54%
1997	3867	23	0.59%
1998	3520	20	0.57%
1999	3509	22	0.63%
2000	3589	17	0.47%
2001	3227	25	0.77%
2002	3077	14	0.45%
2003	2977	17	0.57%
2004	3085	14	0.45%
2005	3066	15	0.49%
2006	2942	20	0.68%
2007	2777	15	0.54%
2008	2430	17	0.70%
2009	1926	12	0.62%
2010	2013	18	0.89%

The consequences of derailment- and non-derailment grade crossing collisions differ considerably (Zhang 2017). In a typical non-derailment grade crossing collision, most casualties are sustained by highway users, and as a result the average number of casualties per incident is fairly low (0.27 per incident) (Figure 1.2a; Table 1.2). For grade crossing collisions that result in derailment, casualties are almost equally split between highway users, railroad passengers, and railroad employees (Figure 1.2b; Table 1.2). As a result, the average number of casualties per incident is almost 7 times greater (1.85 per incident).



(a)



(b)

Figure 1.2: Proportion of casualties sustained by highway users (gray), railroad passengers (orange) and railroad employees (blue) in (a) non-derailment grade crossing collisions and (b) derailment grade crossing collisions (data from Zhang 2017)

Table 1.2: Number of Grade Crossing-Related Casualties by Type (data from Zhang 2017)

	Non-Derailments		Derailments	
	Count	Percent	Count	Percent
Highway User Casualties	17,293	0.90	220	0.30
Rail Passenger Casualties	670	0.04	288	0.39
Rail Employee Casualties	1,148	0.06	231	0.31
Total Casualties	19,111		739	
Total Incidents	70,744		399	
Casualty Rate	0.27		1.85	

For freight trains, if a train does not derail, then in addition to casualties to highway users and train crew, damage to the railroad track can result in lost service time and financial impacts. If the train does derail, there is additional potential for a release if the train is carrying hazardous materials. With increased interest in passenger rail transport and the growth in transportation of hazardous materials such as crude oil, the importance of comprehensive understanding of the risk of grade crossing collisions is more critical than ever.

There has been only limited research on grade crossing-caused derailments. Cherchas et al. (1979, 1982) developed a mathematical model and computer simulation to study the dynamics of rail and highway vehicles in grade crossing collisions, and the potential for derailment based on the train's resulting L/V ratio (the ratio of lateral to vertical force the train applies to the rail). However, Cherchas et al.'s study was directed at a limited set of questions and they focused exclusively on passenger trains, since their goal was to inform decisions about increased passenger train speeds in Canada.

The need to conduct additional research has been recognized by multiple agencies. The risk of grade crossing-caused derailments was identified as one of the 27 risk factors in the Rail Corridor Risk Management System (RCRMS) project initiated by the Railroad Research Foundation (RRF) in 2011(Stephens and Foy, 2008). (The RRF is a non-profit organization funded by the Federal Railroad Administration (FRA), the Department of Homeland Security, and the Association of American Railroads.) The research in this dissertation grew out of an element of the RCRMS project.

The goal of this dissertation is to develop an understanding of the factors that affect derailments due to highway-rail grade crossing collisions, and to develop a statistical model that will enable quantitative assessment of the relative probability that different crossings may cause a derailment. Such a model will enable more informed allocation of safety resources so as to reduce grade-crossing risk. This model could ultimately be integrated into an overarching risk analysis framework that would consider all sources of risk at grade crossings.

1.2. GRADE CROSSING COLLISION EVENT TREE

Simply stated, *risk* can be defined as the *probability* of an event occurring, multiplied by the *consequence* of that event. The probability component can be broken into a series of conditional probabilities to describe the probability of multiple events occurring in sequence. Therefore, in developing a risk model, it is important to understand the chain of events leading to a consequence to identify the conditional probabilities of each event occurrence. The flow chart in Figure 1.3 shows a simplified event chain beginning with a train encountering a grade crossing and ending with a consequence event.

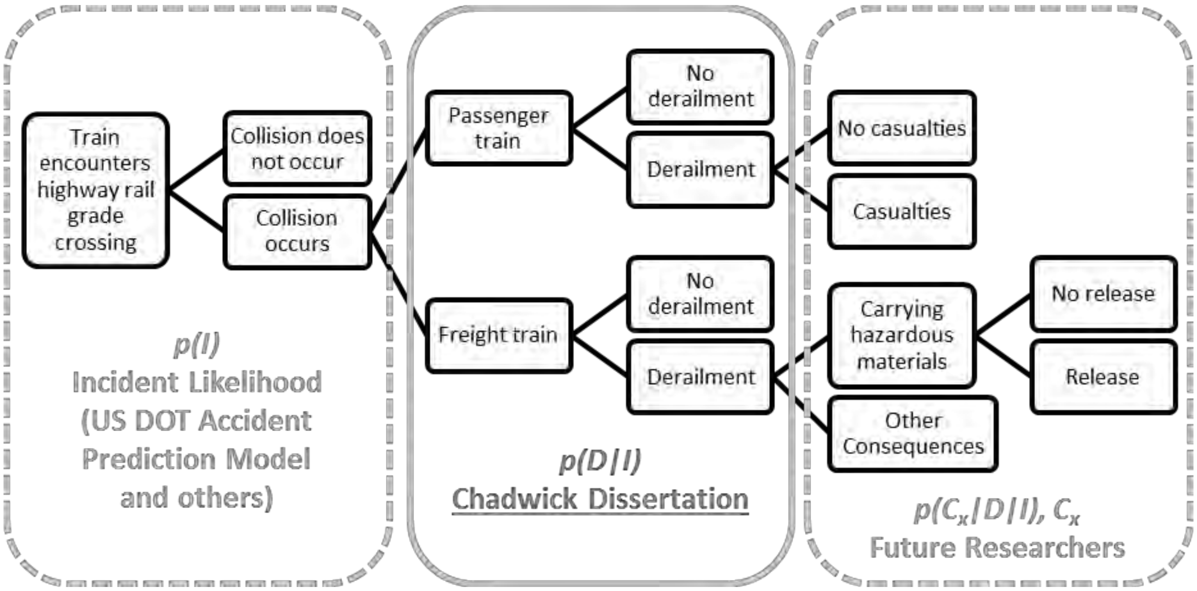


Figure 1.3: Simplified flowchart describing factors affecting risk at highway-rail grade crossings and role of this dissertation.

For a highway-rail grade crossing collision to occur, a train must first encounter a grade crossing. For every grade crossing it encounters, a train has a certain probability of being involved in an incident with a highway vehicle, *p(I)*. This likelihood is dependent on a number of factors, including exposure (traffic on the highway and railroad), visibility, highway speed limit and type of warning device. Extensive work has gone into understanding and quantifying the factors that affect the likelihood of a grade crossing incident. An example of an incident prediction model is the U.S. DOT Accident Prediction Model, which calculates the expected collision frequency at a crossing (FRA, 1987; Ogden and Korve Engineering, 2007).

Given that an incident occurs, there is a probability that a derailment will result, *p(D|I)*. This probability depends on a variety of factors. It is possible that passenger and freight trains may have different derailment likelihoods because of different weight, speed and operating characteristics. Due to their differing cargo, they will also have different consequences. Development of a model for *p(D|I)* is the focus of this dissertation.

Once the derailment has occurred, the probability of each potential consequence can be determined, $p(C_x|DI)$ (Figure 1.3). For freight trains, consequences include employee casualties and railroad financial loss. Another possible consequence of a freight train incident is the release of hazardous materials. Hazardous materials release models have been developed (Saat and Barkan, 2005; Verma and Verter, 2007; Glickman and Erkut, 2007; Liu et al. 2012) that could be combined with the derailment likelihood model to estimate the risk of a hazardous materials release resulting from a grade crossing incident. For passenger trains, consequences include passenger casualties, train crew casualties, and a variety of financial losses. The probability of each consequence occurring, as well as the value of that consequence, should be investigated by future researchers and could be integrated with the derailment likelihood model presented in this dissertation.

Stated as an equation, the overall risk at a crossing is then calculated as:

$$Risk = \sum_{x} p(I) \times p(D|I) \times p(C_x|D,I) \times C_x$$

1.3. RESEARCH FRAMEWORK

To quantify the probability of a grade crossing collision-caused derailment ($p(D|I)$), it is necessary to determine what makes derailment incidents different from non-derailment incidents. A basic hypothesis is that the underlying physical properties of the incidents play a large role. From a simple physics perspective, it is plausible that collisions with more energy would be more likely to result in derailment. Very generally, a derailment involves two bodies (the train and highway vehicle), each with a certain mass and velocity, colliding at a specific angle (Figure

1.4). By quantifying each of these factors, it is hypothetically possible to model how their variation relates to the probability of derailment.

Cherchas et al. (1979, 1982) used a computer model to investigate factors such as the weight and speed of the rail and highway vehicles. The same type of factors motivated the variables investigated and overall statistical approach used in my research.

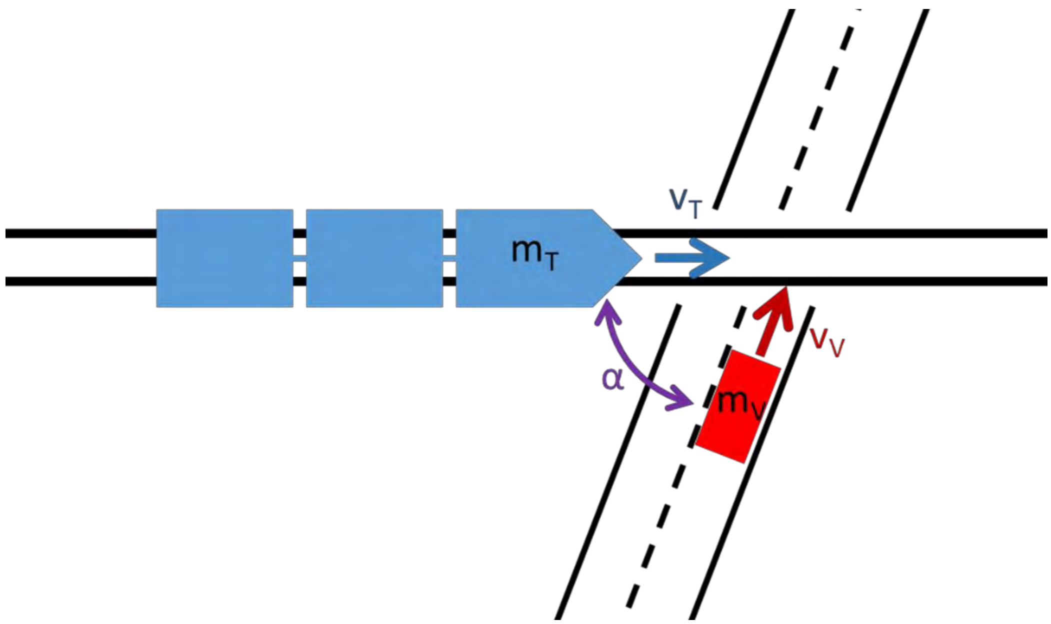


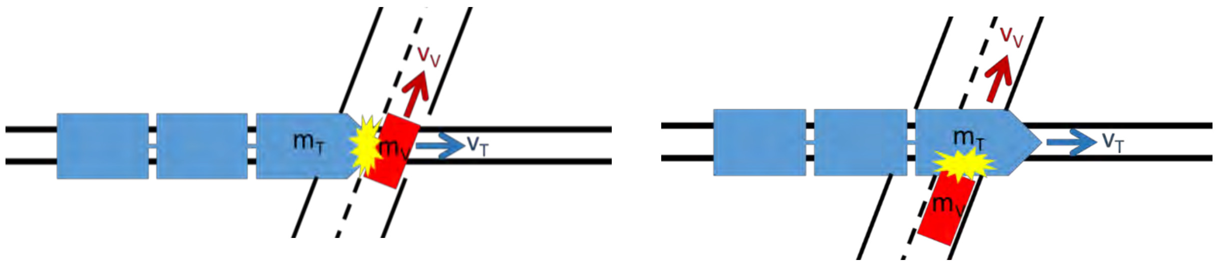
Figure 1.4: Simple free body diagram of the bodies involved in a grade crossing collision.

The variables shown in Figure 1.4 are defined as follows:

v_V: Highway vehicle speed
m_V: Mass of highway vehicle
v_T: Train speed
m_T: Mass of train
α: Angle of collision

An additional factor, *incident type*, accounts for the difference between incidents where the train strikes the vehicle (TSV) and incidents where the vehicle strikes the train (VST) (Figure 1.5). Due to factors such as the interaction between the train's wheels and the rail, the effect of the factors shown in Figure 1.4 are expected to differ significantly for the two incident types.

Cherchas et al. (1982) referred to TSV incidents as “frontal impact” incidents and VST incidents as “side impact” incidents, and developed two different computer models to study their effect on derailment likelihood.



(a) **(b)**

Figure 1.5: Free body diagram for (a) train strikes vehicle (TSV) incidents and (b) vehicle strikes train (VST) incidents.

There are two potential approaches for investigating these factors. The first is to analyze the physical interactions between the highway vehicle, train and track through finite element analysis or other computer modeling techniques. This was the approach used by Cherchas et al. (1979, 1982). The second is to use data collected from highway-rail grade crossing incidents to develop statistical regression models that quantify actual experience to understand the differences between derailment and non-derailment incidents. My dissertation takes the second approach, using twenty years of grade crossing collision data to investigate the relationships between the proposed factors and derailment probability. No previous study has taken advantage of the extensive data collected by FRA to explore this approach. Beside filling a void in this research domain, it has the potential to provide new insights derived from the empirical experience that were not considered in Cherchas et al.’s modeling approach.

I consider each potential factor, first independently through univariate analysis (Chapter 4), then as a component of a multivariate incident-level model (Chapters 5 and 6), and finally in the context of a crossing-specific model (Chapter 7). This leads to the development of the grade crossing incident derailment likelihood calculator (Chapter 8), a Microsoft Excel workbook that can be used to estimate the conditional probability of derailment for a grade crossing based on its specific characteristics. I then describe a case study illustrating use of the calculator and discuss how this model could affect grade crossing upgrade prioritization (Chapter 9). I conclude with a summary of future data and research needs that my work has uncovered (Chapter 10).

1.4. DISSERTATION APPLICATIONS

The results of my research serve three purposes. First, through development of the $p(D|I)$ model I shed light on a variety of underlying factors in grade crossing collision caused derailment likelihood. The basic factors involved in these events have been suggested anecdotally by railroad employees who investigate derailments; however, the factors have not previously been systematically investigated. An understanding of these factors might also lead to techniques for potential derailment avoidance.

Second, my model can serve as a ranking tool for transportation agencies that are trying to determine which grade crossings to upgrade with available funds. If an agency uses a model such as the US DOT Accident Prediction Model, multiple crossings may have similar incident likelihoods, but there may not be sufficient resources to upgrade them all at the same time. The derailment consequence model could be used to further refine the prioritization of upgrades for these crossings.

The third purpose of my research is to put in perspective the relative likelihood of catastrophic, grade-crossing-collision-caused derailments compared to other sources of railroad and highway risk. Preliminary evaluation of the expected number of lives lost as a result of grade-crossing-caused-derailments, compared to the expected lives lost due to grade crossing collisions generally, shows that for most crossings, derailments should not be the primary source of concern. This research can lead to a cost-benefit analysis of different safety improvement programs that can be undertaken by railroads and communities, to determine how resources can be most effectively used to reduce casualties.

Since I began work on this project several years ago, I have had numerous opportunities to present my research to members of the railway industry, government officials and academic researchers. Representatives from all three groups have expressed interest in applying the results of my research in their work. I believe this dissertation advances the state of the art in grade crossing safety and risk analysis, and provides useful insights and tools to all three groups.

1.5. DEFINITIONS / GLOSSARY

Throughout this dissertation, I use consistent terminology to describe the objects involved in a grade crossing incident. This section outlines these definitions.

A grade crossing incident is defined as any collision between a rail consist and a highway user at a grade crossing[1]. A grade crossing derailment is defined as any grade crossing incident where one or more cars or locomotives were derailed because of the incident.

A rail vehicle is here defined as any equipment that is traveling on railroad track. Rail vehicle refers to both locomotives and railcars, regardless of whether they are passenger or freight equipment.

A highway vehicle is any motorized vehicle traveling on the roadway system. Unless specified, it refers to both large highway vehicles (semi-trucks and straight trucks) as well as small highway vehicles (personal automobiles, pickup trucks and vans). It does not include non-motorized vehicles such as bicycles, or pedestrians.

Incident type is a variable used to differentiate between incidents where the train struck the vehicle (TSV) and the vehicle struck the train (VST). This is not a standard terminology, and other sources refer to these incidents by different names. In the HRA database, they are referred to as "rail equipment struck highway user" (TYPACC = 1) and "rail equipment struck by highway user" (TYPACC = 2).

FRA track class is a rating system used by the Federal Railroad Administration to assess track condition. Track class for a rail segment is determined based on the maximum speed at

---

[1] The FRA defines a grade crossing accident as "any collision, derailment, fire, explosion, act of God, or other event involving operation of railroad on-track equipment (standing or moving) that results in damages greater than the current reporting threshold to railroad on-track equipment, signals, track, track structures, and roadbed." A grade crossing incident is defined as "any event involving the movement of on-track equipment that results in a reportable casualty but does not cause reportable damage above the current threshold established for train accidents" (FRA, 2011b).

which a railroad wishes to operate trains. In turn, this determines the standard to which the track must be maintained. The maximum speed for a track class differs for freight and passenger trains (Table 1.3).

Table 1.3: FRA Track Class and Speed Limits

Track Class	Maximum Speed (Freight)	Maximum Speed (Passenger)
X (Excepted)	10	--[1]
1	10	15
2	25	30
3	40	60
4	60	80
5	80	90
6	110	110

[1]Revenue passenger trains are not allowed to operate on excepted track (49 CFR 213.9, 2011).

Timetable speed is effectively the "speed limit" for trains on a segment of track. It is related to FRA track class, in that the maximum speed permitted based on track class is the upper limit of the timetable speed. However, track segments frequently have a lower timetable speed than that permitted by their FRA track class due to civil speed restrictions and other factors.

Grade crossing warning devices are any sign or signal added to the roadway at or near a grade crossing to help alert motorists to the presence of a grade crossing. A distinction is typically made between "passive" and "active" devices. Passive devices do not change state when a train approaches the crossing, and therefore alert motorists to the presence of a railroad track but not the presence of a train. The most common passive device is the crossbuck (or railroad crossing sign). In contrast, active devices employ technology that detects a train approaching a crossing and activates systems to alert drivers (typically a set distance or time before the train reaches the crossing). For example, flashing lights will illuminate when a train approaches. The most common active devices are flashing lights (or flashers), bells, and gates. In this dissertation, I further differentiate between “gates” and “other active” devices, because gates provide a physical warning blocking the roadway in front of motorists entering the crossing, whereas flashers or bells only provide visual and audible cues.

CHAPTER 2: LITERATURE REVIEW

This chapter provides an overview of literature related to grade crossing safety. I begin with background information regarding regulation of highway-rail grade crossings in the U.S., then discuss incident likelihood models and information on the statistical techniques employed in later chapters. I conclude with some information on alternative grade crossing warning devices and human factor aspects of crossing safety.

2.1. GRADE CROSSINGS AND LEGISLATION IN THE UNITED STATES

Highway-rail grade crossings in the U.S. are regulated by the government through a number of federal and state agencies. Much of the funding for grade crossing warning device upgrades comes from the federal government. This funding has been distributed through a series of acts and programs over the years, including the 1991 Intermodal Surface Transportation Efficiency Act (ISTEA), and the 2005 Safe, Accountable, Flexible, Efficient Transportation Equity Act – A Legacy for Users (SAFETEA-LU). Other funding is provided by local transportation agencies. Additionally, railroads often contribute to crossing upgrade projects, either financially or through provision of labor (Ogden and Korve Engineering, 2007).

In general, railroads are responsible for design of a grade crossing on tracks, including crossing surfaces and, if present, whatever active warning system that may interface with the track, since these systems require specialized maintenance expertise (Ogden and Korve Engineering, 2007). Railroads are also required to install and maintain crossbuck signs (the minimum level of warning device required by law at public grade crossings). Beyond the railroad right-of-way, responsibility for advance crossing warning signs and interconnected highway signals falls to the local highway agency. Standards for traffic control signs at grade crossings are maintained by the Federal Highway Administration (FHWA) and codified in the Manual on Uniform Traffic Control Devices (MUTCD) (FHWA, 2012). Conditions at grade crossings are legally required to meet "accepted standards and practices", especially those outlined in MUTCD (Ogden and Korve Engineering, 2007). Beyond this, all states are required to have a highway safety improvement plan (HSIP) that includes the ability to prioritize grade crossing improvements in the state.

As train speeds have increased, special grade crossing regulations have been developed governing high-speed sections of track (Table 2.1). The Federal Railroad Administration (FRA) has issued regulations requiring complete grade separation for railroad operations with FRA track class 8 or higher, meaning operating speed in excess of 125 mph (referred to as "high speed rail", or HSR). For "higher-speed rail" (HrSR) operations between 110 and 125 mph on class 7 track, grade crossings are not recommended but may still be used with certain extra protections. The FRA requires active warning devices on all crossings on class 6 track. For track class 5 and lower, there are no specific requirements beyond the presence of a crossbuck at public crossings. Outside the Northeast Corridor most track is FRA class 5 or lower, although classes 6 and 7 are slowly expanding as higher speed passenger services are implemented.

Table 2.1: Summary of federal regulation related to grade-crossings (CFR, 2012)

Maximum Passenger Train Speed	79-110 mph (127-177 kph)	110-125 mph (177-201 kph)	> 125 mph (201 kph)
Track Class	6	7	8-9+
Grade Crossing Protection Type	Active[1]	Warning/Barrier[2] with FRA Approval	Grade Separate or Close

[1] FRA recommends but does not require sealed corridor treatment for all grade crossings with train speeds in excess of 79 mph (127 kph)
[2] Barrier system must physically prevent incursion of motor vehicle into right-of-way

The most economical approach to eliminating the risk of collisions at a crossing is to close it; however, local communities are often opposed to this because of a real or perceived loss of convenience. Additional concerns may be raised about increased emergency service response time and reduced access to schools and other strategic places. If a crossing cannot be closed, other approaches may be considered including grade separation and upgraded warning and protection devices.

2.2. GRADE CROSSING ACCIDENT PREDICTION

Many methods of modeling collision likelihood at grade crossings have been developed, mainly with the goal of understanding the risk posed to highway users by trains. These models have traditionally been used to decide how funds for highway-rail grade crossing improvements should be allocated. However, the collision rates predicted by these models can also be used to quantify risk to the train, its passengers, or its cargo.

Faghri and Demetsky (1986) categorized collision likelihood models into two groups: relative formulas and absolute formulas. Relative formulas use crossing data to rank the relative hazards at each crossing so that improvements can be prioritized from most dangerous to least dangerous crossings. Absolute formulas predict the number of collisions expected to occur at each crossing over a certain time period (i.e. a rate), thereby allowing estimation of the casualties prevented if a crossing is upgraded.

Formulas to rank grade crossings by their collision risk were developed at least as early as the 1940s (Austin and Carson, 2002; Faghri and Demetsky, 1986). Many of these formulas are still used by state departments of transportation. The four formulas presented here are the New Hampshire model, the Peabody-Dimmick (or Bureau of Public Roads) formula, the NCHRP

Report 50 Hazard Index, and the U.S. DOT Accident Prediction Model. They are presented in order from least to most complex.

The New Hampshire model is a relative formula that can be used to rank the importance of crossing upgrades (Austin and Carson, 2002; Faghri and Demetsky, 1986). It has been widely used across the country, either in its original form or with various modifications. Analysis has shown that the New Hampshire hazard index ranks crossings similarly to more complex formulas, but it is limited in that it does not predict the expected number of collisions. The hazard index formula is:

$$Hazard\ Index = V\ T\ P_f \tag{1}$$

Where:

V = average 24-hour (highway) traffic volume
T = average 24-hour train volume
P_f = protection factor (0.1 for gates; 0.6 for flashing lights; 1.0 for signs only)

The Peabody-Dimmick formula (also called the Bureau of Public Roads formula), developed in 1941, is an absolute formula that predicts the number of accidents at a crossing over a period of five years (Austin and Carson, 2002; Faghri and Demetsky, 1986). The five-year accident prediction formula is:

$$A_5 = 1.28\frac{(V^{0.170})(T^{0.151})}{P^{0.171}} + K \tag{2}$$

Where:

A_5 = expected number of accidents in five years
V = annual average daily traffic (AADT)
T = average daily train traffic
P = protection coefficient
K = additional parameter (smoothing factor)

The NCHRP Report 50 Hazard Index is an absolute formula developed in 1968 by Andrew Voorhees and Associates (Austin and Carson, 2002; Faghri and Demetsky, 1986). It can be expressed as a formula, but is more commonly determined from a series of charts and tables that allow the user to calculate the expected yearly accident rate. It is dependent on factors such as AADT, number of trains per day, type of warning device in use, the geographic location of the crossing, and geometric aspects of the crossing.

Today, the most commonly used model is the U.S. DOT Accident Prediction Model (Austin and Carson, 2002; Faghri and Demetsky, 1986; Ogden and Korve Engineering, 2007). First developed in the early 1980s, this absolute formula uses nonlinear multiple regression techniques on a wide variety of factors, including highway type and train traffic, to predict the expected yearly number of collisions at a crossing. The general expression of the formula is:

$$a = K \times EI \times MT \times DT \times HP \times MS \times HT \times HL \tag{3}$$

$$B = \frac{T_0}{T_0 + T}(a) + \frac{T}{T_0 + T}\left(\frac{N}{T}\right) \tag{4}$$

$$A = \begin{cases} 0.7159B & \text{For passive devices} \\ 0.5292B & \text{For flashing lights} \\ 0.4921B & \text{For gates} \end{cases} \tag{5}$$

Where:

a = initial collision prediction, collisions per year at the crossing
K = formula constant
EI = factor for exposure index based on product of highway and train traffic
MT = factor for number of main tracks
DT = factor for number of through trains per day during daylight
HP = factor for highway paved (yes or no)
MS = factor for maximum timetable speed
HT = factor for highway type
HL = factor for number of highway lanes
B = adjusted accident frequency value
T_0 = formula weighting factor; $= 1.0/(0.05 + a)$
T = the number of years under study
N = number of observed accidents in T years at a crossing
A = normalized accident frequency value

A table provides each of the factors, for crossings with passive devices, flashing lights, and gates. The U.S. DOT also provides a procedure for using this formula to determine grade crossing upgrade resource allocation (FRA, 1987).

Faghri and Demetsky (1986) tested four absolute formulas and found that the U.S. DOT formula most accurately predicted the number of collisions occurring at grade crossings in Virginia for the five-year period of study. They recommended that the Virginia DOT use the U.S. DOT formula in combination with site visits to evaluate the importance of grade crossing upgrades.

However, are some concerns remain about the accuracy of the U.S. DOT model. It is based on data from the entire U.S. so it may not account for regional differences. As a result, some states have developed specialized formulas using more detailed state-specific data. For example, Benekohal and Elzohairy (2001) examined ten years of highway-rail grade crossing collisions in Illinois. They found that the U.S. DOT formula selected only 89 of the top 200 grade crossings with collisions for upgrade, and did not reliably identify the most dangerous crossings. They developed a regression model, the Illinois Hazard Index, that suggested a higher percentage of crossings with collisions for improvement; it also selected locations with higher crash rates compared to other equations.

Another concern about the accuracy of the U.S. DOT formula is that crossing conditions and warning/protection technologies may have changed since its development. Austin and Carson (2002) showed that the normalizing coefficients used in Equation 5 to account for the difference between the model's predicted and observed values have been steadily declining over time; that is to say, the model's prediction accuracy has diminished, and the normalizing coefficients have been adjusted to compensate. Austin and Carson proposed that the formula's accuracy could be improved if it were re-evaluated using present-day data. However, they also considered the complexity of the U.S. DOT's three-part formula to be problematic, since it is difficult to interpret and prioritize the effects of changing various parameters. To address this concern, Austin and Carson developed an alternate model using negative binomial regression. This model identified many of the same significant variables as the U.S. DOT formula, but as it was developed using only collision data at public crossings in six states, further testing would be needed to see if it is more broadly applicable.

Chaudhary et al. (2011) compared the performance of the U.S. DOT Accident Prediction Formula to that of the Transport Canada Accident Model to see which would more effectively identify "hot spots" (high-risk areas) on a network in California. Overall, the U.S. DOT model more closely predicted the annual number of accidents occurring at a crossing. However, in cases where the crossing had an accident history, the Transport Canada model was more accurate. Chaudhary et al. suggested adapting the Transport Canada model to U.S. crossing data and using it to rank the most dangerous crossings.

A variety of statistical models have been developed with the goal of improving the accuracy of collision frequency prediction of the earlier models. The following five papers are a sample of those using a variety of advanced statistical methods including Poisson regression, negative binomial regression, gamma probability models, and Bayesian analysis.

Saccomanno et al. (2004) developed models for identifying highway-rail grade crossing black-spots in Canada, where a "black-spot" is an area that is unusually prone to incidents. They performed both a Poisson regression and a negative binomial (NB) regression on the data, considering 11 variables. For each regression method, three separate models were developed – one for passive crossings, one for crossings with flashing lights, and one for crossings with gates.

They found that the NB models performed better. Significant factors varied by crossing treatment. For passive crossings, train speed and exposure were the only significant factors. For crossings with flashing lights, surface width was also found to be significant. For crossings with gates, road speed, number of tracks and exposure were found to be significant. They also found that crossings with the highest collision frequency were in urban areas, probably because the exposure factor is likely to be higher in urban areas than in rural areas.

Other frequency models have been developed abroad. South Korea evaluated the effectiveness of the U.S. DOT Accident Prediction Formula for predicting accidents at Korean grade crossings (Oh et al., 2006). They found that the U.S. DOT formula did not accurately predict collision rates in South Korea. This may be because, unlike the U.S, all grade crossings in South Korea are equipped with gates. They then developed a gamma probability model using Korean accident data. Collisions were observed to increase with highway traffic volume, train volume, proximity to commercial areas, distance of train detector from crossing, and time between activation of warning signals and gates. An interesting aspect of Oh et al.'s work is that grade crossing warning device type, which is a critical factor in most other collision prediction models, is eliminated from the analysis because all crossings in South Korea have the same warning device. This illustrates that some factors, such as proximity to commercial areas, might affect collision rates in the U.S. but are not considered in U.S. models.

Washington and Oh (2006) and Saccomanno et al. (2007) both used Bayesian methodology to assess the effectiveness under uncertainty of different grade crossing treatments at reducing collision rates. Washington and Oh analyzed 18 grade crossing treatments. For each treatment, a survey of past research findings was conducted to determine a Bayesian prior density. Next, a panel of experts was presented with a random sample of collisions that had occurred at South Korean grade crossings, and were asked to evaluate how each treatment would have affected the occurrence of each collision. This information was aggregated into a best-estimate “current” accident modification factor (AMF). Bayes’ theorem was then used to combine the experts’ AMF with prior knowledge to obtain “posterior” intervals of AMFs. Applying this methodology, Washington and Oh determined that the three highway treatments that most effectively reduce crashes are in-vehicle warning systems, crossing obstacle detection systems, and constant warning time device activation systems. As the authors point out, the rankings do not consider cost; however, this is a critical factor to consider when upgrading a crossing and should be included in further analysis. Additionally, their ranking of treatments does not account for variation among crossing locations. Saccomanno et al. (2007) took a different approach in their analysis. Their model can be calibrated to a specific crossing and provides a statistical distribution of the expected change in collision likelihood for a given treatment. Since grade crossing collisions are random events, and there is a lack of before-and-after studies for many grade crossing treatments, engineers must make decisions about grade crossing upgrades under uncertainty. By considering a range instead of an absolute value for the reduction of risk, engineers will be able to better evaluate the cost-benefit ratio for a given crossing upgrade.

Mok and Savage (2005) took a different approach to analyzing collision rates at grade crossings. They observed that the number of collisions and fatalities at grade crossings in the U.S. has decreased significantly over the past 30 years, despite an increase in both train and highway traffic. Their analysis showed that about 70% of the decrease could be attributed to human-factor related aspects (such as educational programs like Operation Lifesaver, and the requirement of ditch/crossing lights on locomotives), and 30% could be attributed to the installation of gates and/or flashing lights and crossing closures. This result suggests that collision prediction models rightly attribute high importance to the type of crossing warning device in use, but should also consider human factor aspects.

Overall, a variety of models for the prediction of collision rates at highway-rail grade crossings have been developed. In general, there appear to be trade-offs between ease-of-use, accuracy, and specificity. The most accurate models are generally developed for a small set of all collisions – for example, considering data from only one state in the U.S. This may mean that each individual state would need to create its own model using its own data – a task that may be worthwhile to more reliably identify the most dangerous crossings. Additionally, the more accurate models may be more difficult to use for decision-makers who do not have a statistical background. This problem could likely be avoided if a straight-forward procedural document were developed, as was done for the U.S. DOT Accident Prediction Formula.

2.3. ACCIDENT SEVERITY

In terms of overall risk, the frequency of an event occurring is half the equation; the other half is the consequence of that event. Most previous research has focused on the risk grade crossings pose to highway vehicle users. These represent a large percentage of railroad-related fatalities (FRA, 2017c). However, it is also important to study the impact collisions could have on train crew and passenger injuries and fatalities, as well as hazardous materials. This section presents statistical models of severity, as well as crashworthiness research, that seeks to understand the physical forces involved in a collision and how to mitigate them.

2.3.1. Statistical modeling of severity

The U.S. DOT developed equations for predicting the probability of an injury accident given an accident and the probability of a fatal accident given an accident (Ogden and Korve Engineering, 2007). The expressions for these probability equations are:

$$P(IA|A) = \frac{1 - P(FA|A)}{1 + CI \times MS \times TK \times UR} \tag{6}$$

$$P(FA|A) = \frac{1}{1 + CF \times MS \times TT \times TS \times UR} \tag{7}$$

Where *CI* = formula constant = 4,280

CF = formula constant = 695

MS = factor for maximum timetable train speed

TK = factor for number of tracks

TT = factor for through trains per day

TS = factor for switch trains per day

UR = factor for urban or rural crossing

The FHWA Grade Crossing Handbook (Ogden and Korve, 2007) provides tables listing the value of each factor based on the initial prediction of the base collision likelihood model, as well as the number of observed accidents at the crossing over the past one to five years. The results of the injury and fatality likelihood models can be used in conjunction with the accident likelihood model to prioritize the most dangerous crossings.

Limited research has been conducted quantifying the consequences of grade crossing collisions from the rail perspective. Cherchas et al. (1979, 1982) developed a computer simulation tool to analyze the dynamics of highway-rail grade crossing collisions, specifically the relationship between derailment likelihood and train speed in response to proposed higher-speed passenger train operations in Canada. The simulation also enabled them to investigate the effects of highway vehicle size and speed on derailment likelihood. Their simulation suggested that train derailment likelihood should increase with highway vehicle mass and speed. The model also indicated that derailment likelihood was a function of train speed; however, this effect changed based on the exact location of the highway vehicle's impact on the rail vehicle (for example, whether the highway vehicle struck the front of the locomotive or its side). The simulation used the ratio of railhead lateral force to vertical wheel load (L/V ratio, or derailment coefficient) as the criteria for derailment. The results were calibrated using experiments conducted by Japan National Railways in the early 1970s to show that they reflected real-world scenarios. Cherchas' study only considered passenger train consists, because the motivation for their study was to inform safety and risk aspects of possible increased passenger train service in Canada. Consequently, the results cannot be directly applied to freight trains.

2.3.2. Crashworthiness

U.S. researchers have conducted extensive studies of rail passenger car crashworthiness. This research has focused primarily on American commuter and intercity rail cars traveling at speeds below 100 mph.

Simons and Kirkpatrick (1998) developed a finite element model of a theoretical generic U.S. high speed passenger train and then used it to understand the safety risks posed to passengers. The train consist was tested in seven different crash scenarios, mimicking head-on collisions with various objects at various speeds. For each scenario, the expected number of casualties was predicted based on primary and secondary impact data. Collisions at speeds of 60 and 100 mph with a 50-ton object show what would happen if a typical higher-speed train collided with a large highway vehicle such as a tractor-semitrailer. Simons and Kirkpatrick demonstrated the potential for severe casualty levels in grade crossing collisions.

A 1998 collision in Portage, Indiana between a commuter train and a tractor-trailer carrying steel coils led to new regulations addressing passenger rail car structural design. Full-scale collision testing of the new passenger cars was conducted to compare their performance to the pre-1999 car design. Jacobsen et al. (2003) tested the crash performance of the two car designs by colliding them with a steel coil truck to imitate the Portage incident. They found that the 1990's cab-car end structure deformed more than 20 inches (50.8 cm) longitudinally, resulting in loss of operator survival space, whereas the new design deformed only 8 inches (20.3 cm) and preserved survival space. Martinez et al. (2003) developed a computer model to predict crushing behavior in the cab car. They validated their model with the full-scale collision test, and found that the model accurately predicted crush patterns. Samavedam and Kasturi (2011) performed the same full-scale test at higher speeds to validate their finite element model (FEM) of train collisions. The model closely predicted the overall damage to the locomotive, as well as predicting the intrusion into operator survival space.

In the wake of the 2005 Glendale, California collision between a Metrolink commuter train and an SUV, in which 11 people were killed, the FRA released a report on the safety of push-pull and multiple-unit locomotive passenger rail operations (FRA, 2006). This report sought to understand the relative crashworthiness of multiple-unit electric (EMU) cars, cab-car leading trains ("push mode"), and conventional locomotive-led trains ("pull mode"). Analysis of 20 years of data showed that, while locomotive-powered trains operated in the push mode had a slightly greater number of fatalities and tendency to derail than those operated in pull mode, the differences were not statistically significant. EMUs were shown to have a superior safety record, with the lowest fatality rate of all passenger transportation modes (including air travel).

Also in response to the Glendale collision, Metrolink worked with the FRA, the Federal Transit Administration, and the American Public Transportation Association to develop a performance-based technical specification for railroad passenger car crashworthiness, focused on crash energy management (CEM). This work resulted in performance specifications for the overall train consist; for the cab and passenger-carrying cars; and for mechanical components such as couplers (Tyrell et al., 2006). CEM trains are designed to deform in a controlled way during a collision, collapsing unoccupied areas to absorb energy and preserving survival space in the occupied areas. Tyrell and Perlman (2003) compared the crashworthy (or survivable) speeds of CEM and conventional trains, in both train-to-train collisions and highway-rail grade crossing collisions. They found that passengers in CEM trains could experience a much higher primary collision speed and survive, even though their secondary impact velocity would be slightly greater than in a conventional train.

2.4. ALTERNATIVE GRADE CROSSING WARNING STRATEGIES

The formulas for highway-rail grade crossing collision likelihood and severity discussed above account for existing grade crossing warning devices and strategies. Typical levels of protection in the U.S. are passive crossings with crossbucks or stop signs, active crossings with flashing lights and bells, and gated crossings that include flashers and bells. Crossings with a history of collisions can be upgraded to more restrictive warning devices. If a crossing has a full gate treatment and still experiences a high collision rate, railroads and local departments of transportation may work together to close the crossing.

Ideally, rail lines on which passenger trains or trains carrying hazardous materials operate whould be completely grade separated; however, due to cost and other factors it may be infeasible to grade-separate an entire line. New ideas on grade crossing protection include augmented passive systems and sealed corridors. Nelson (2010) reviews numerous strategies to reduce grade crossing risk currently in use around the world. These include closures and consolidation, upgraded lights and gates, and alternative technologies such as in-pavement flashers. The goal in the U.S. is to develop a strategy that balances cost with risk reduction.

2.4.1. Sealed Corridors

The sealed corridor concept has been developed as a way to upgrade conventional rail lines to carry higher-speed passenger trains. For trains operating in the 110-125 mph range, grade separation is suggested but not required; instead, the FRA requires crossings to have approved barrier systems that can prevent highway vehicle incursion onto the right of way. Obstacle detection systems to alert the train if a vehicle becomes stuck on the tracks are also recommended. These requirements and appropriate technologies for use in achieving these goals were summarized by FRA (2009a).

The state of North Carolina (NC) was the first to make aggressive use of the sealed corridor concept (Bien-Aime, 2009; FRA, 2009a, 2009b). The NC DOT Sealed Corridor is part of the Southeast High Speed Rail (SEHSR) Corridor and included 216 grade crossings, 44 of which were private crossings. Between 1987 and 2004, this section experienced 282 collisions, resulting in 74 injuries and 55 fatalities to motorists. The program has consolidated as many grade crossings as possible and upgraded the rest to include self-monitoring four-quadrant gates, long-arm gates, and/or traffic channelization devices. NC DOT projects that 19 lives were saved between 2004 and 2009 by implementing the sealed corridor concept due to a sharp decrease in the number of grade crossing collisions (Bien-Aime, 2009; FRA, 2009c).

Illinois DOT (IDOT) has been upgrading sections of track for 110 mph operation between Chicago and St. Louis using a sealed corridor approach (Hellman and Ngamdung, 2009). The route between Chicago and Springfield, IL had 311 grade crossings, of which 68 were proposed for closure. However, only 10 crossings were ultimately closed due to strong opposition from impacted communities. Of the remaining crossings, 69 were equipped with four-quadrant gates and vehicle detection systems. The results in North Carolina suggest that this approach will reduce collisions and fatalities along the route, even with higher speed passenger trains.

2.4.2. Obstacle Detection

Glover (2009) summarizes the goal of obstacle detection as “identifying the presence of a vehicle or person on the crossing as the train approaches and communicating this to the train driver in time for him or her to stop before reaching it.” Obstacle detection technology may mitigate grade crossing risk; however, a substantial challenge is that these systems provide relatively little time to react to an intrusion and stop a train. Glover suggests that there may only be a limited reduction in the severity of an incident because a train may still collide with the obstructing highway vehicle, though at reduced speed. Additionally, Glover cites concerns that these systems could be less reliable than traditional gated crossings; since the devices are fail-safe, an error in the detection system could result in a “false-positive” that closes the crossing gates when no train is approaching. If highway users become accustomed to higher error rates, they may erroneously assume the crossing is out of service, when in fact the gates have been activated by the presence of a train. If they attempt to circumvent the gates, a collision could occur.

Hall (2007) suggests that there are benefits to obstacle detection even if the system is not entirely effective in preventing collisions. Advance warning of a track obstruction could allow the train to decelerate sufficiently to reduce the likelihood of train passenger deaths, especially when combined with more crashworthy passenger train designs. Additionally, Hall states that obstacle detection systems will have the greatest benefit when information can be communicated directly between the grade crossing and an approaching locomotive (such as might be possible with some forms of Positive Train Control (PTC)).

Obstacle detection systems are being used both domestically and abroad. On the line between Chicago and St. Louis, which is being upgraded for 110 mph operation, IDOT and Union Pacific (UP) use a detection system consisting of an inductive loop embedded in the pavement on either side of the track. It is capable of detecting the presence of a vehicle within the crossing gates and alerting the approaching train through in-cab signaling (Hellman and Ngamdung, 2009). This system could also be integrated with a PTC-equipped train consist.

The system usually operates in “dynamic” mode, meaning the exit gates function based on the presence of highway vehicles within the grade crossing. However, in the fail-safe condition, it operates in a “timed” mode that closes the exit gates after a specified amount of time. The FRA and the Volpe Center conducted tests of this equipment to verify its reliability. They found that the average total delay caused by malfunction of this gate equipment to the five higher-speed passenger round-trips that occur daily was approximately 38.5 minutes. They also found that this equipment had a “minimal impact on the frequency and duration of grade crossing malfunctions” (Hellman and Ngamdung, 2009).

2.4.3. Traffic Channelization

Traffic channelization devices direct or separate traffic flow. In the context of highway-rail grade crossings, these devices are intended to prevent drivers from using a grade crossing in an unsafe manner by confining them to controlled lanes (FRA, 2010). An example is a raised median. Research has suggested that channelization discourages risky driving behavior around grade crossings, such as "zig-zagging" past closed gates. Several states have already begun to employ channelization in an effort to improve grade crossing safety.

2.4.4. Low-Cost Grade Crossing Warning Devices

An emerging trend in grade crossing warning devices is the development of low-cost systems that provide a level of safety comparable to conventional devices. These systems generally cost between 5% and 30% of conventional technologies and often rely on wireless communications and solar power (FRA, 2011c). Wullems (2011) summarizes the state-of-the-art of this technology and considers its potential for large rural networks such as on the Australian rail network. Hellman and Ngamdung (2010) described several low-cost warning devices that satisfy the FRA's minimum performance requirements for grade crossing warning devices. They emphasized the importance of reducing annual maintenance costs, not just installation costs.

While low-cost grade crossing warning devices may be interesting from a cost-efficiency point of view, it is unlikely that the devices currently on the market will be used for higher-speed, shared-corridor applications in the U.S. They do not incorporate gates and instead rely on augmented passive systems (adding lights or advance warnings to areas around crossings). Additionally, there are significant legal concerns stemming from public perception of the devices, fail-safe requirements and liability to both the public and private sector (Hellman and

Ngamdung, 2010; Wullems, 2011). However, the concept will likely continue to develop and may expand to include gate technology.

2.5. HUMAN FACTORS AND DRIVER BEHAVIOR

Understanding driver behavior and identifying human factors that cause accidents at highway grade crossings can contribute to development of better accident-prevention strategies. This section provides an overview of literature relating to human factors and driver behavior at grade crossings, as they might pertain to shared corridor operations. For an excellent in-depth review of all literature on driver behavior at grade crossings, see Yeh and Multer (2008).

Caird et al. (2002) developed a taxonomy of human factor accident contributors to highway-rail grade crossing accidents. The taxonomy groups common accident contributors into six categories: unsafe actions, individual differences, train visibility, passive signs and markings, active warning systems, and physical constraints.

People react differently to warning signs at grade crossings. Several studies have been conducted with the goal of identifying the source of this variation (Lenné et al., 2011; Jeng, 2005; Tey et al., 2011a, 2011b; Caird et al., 2002). On average, males are involved in crossing fatalities more than females (Raub, 2007; Caird et al. 2002). The age group of 26 to 64 accounted for the most fatalities (Caird et al., 2002); however, this age group drives the most and thus has the greatest exposure (Evans, 1991). Different age ranges within this group might have different results. Taylor (2008) stated that 16- to 25-year-old drivers were the group most at risk at grade crossings because they were most likely to engage in risky crossing behavior.

In response to warning signs at grade crossings, drivers showed lower compliance rates at passive crossings than at active crossings (Lenné et al., 2011; Tey et al., 2011a, 2011b).

Additional warnings, especially the addition of active warning devices, should result in increased crossing compliance. However, due to cost, it is not feasible to update all passive signs to active warning systems. Alternative ways of augmenting passive crossings are being studied (Cairney, 2003; Tey et al., 2011b; Wullems, 2011; also see Section 2.4.4). Caird et al. (2002) summarized the effectiveness and cost of different countermeasures at grade crossings (Table 2.2).

Table 2.2: Caird et al. (2002) summary of effectiveness and cost of countermeasures

Countermeasure	Effectiveness	Cost	References
Stop signs at passive crossings	Unknown	$1.2 to $2 K (US)	NTSB (1998)
Intersection lighting	52% reduction in nighttime accidents over no lighting	Unknown	Walker and Roberts (1975)
Flashing lights	64% reduction in accidents over crossbucks alone; 84% reduction in injuries over crossbucks; 83% reduction in deaths over crossbucks	$20 to $30 K (U.S.) in 1988	Schulte (1975) Morrissey (1980)
Lights and gates (2) + Flashing lights	88% reduction in accidents over crossbucks alone; 93% reduction in injuries over crossbucks; 100% reduction in deaths over crossbucks 44% reduction in accidents over flashing lights alone	$150 K (U.S.)	NTSB (1998) Schulte (1975) Morrissey (1980) Hauer and Persaud (1986)
Median barriers	80% reduction in violations over 2-gate system	$10 K (U.S.)	Carroll and Haines (2002a)
Long arm gates (3/4 of roadway covered)	67 to 84% reduction in violations over 2-gate system	Unknown	Carroll and Haines (2002a)
4-quadrant gate systems	82% reduction in violations over 2-gate system	$125 K (U.S.) from standard gates; $250 K (U.S.) from passive crossing	Carroll and Haines (2002a), Hellman and Carroll (2002)
4-quadrant gate system + median barriers	92% reduction in violations over 2-gate system	$135 K (U.S.)	Carroll and Haines (2002a)
Crossing closure	100% reduction in violations, accidents, injuries and deaths	$15 K (U.S.)	Carroll and Haines (2002a); NTSB (1998)
Photo/video enforcement	34 to 94% reduction in violations	$40 to $70 K per installation (U.S.)	Carroll and Haines (2002b)
In-Vehicle Crossing Safety Advisory Warning Systems (ICSAWS)	Unknown	$5 to $10 K (U.S.) per crossing + $50 to $250 per receiver	NTSB (1998)

Caird et al. (2002) and Sussman and Raslear (2007) classified the primary reasons for accidents at grade crossings as intentional, distraction-caused or other (visibility issues or driver confusion) for both passive and active grade crossings. Each of these requires a different approach to reducing incident occurrence.

A system for addressing such problems is referred to as the “Three Es”: engineering, education, and enforcement (Jeng, 2005; Sussman and Raslear, 2007). “Engineering” involves using better devices or systems to alert people to the presence of a grade crossing, or to prevent them from entering it. “Education” aims to increase public awareness of the hazards of train movements, as well as reduce dangerous behaviors. “Enforcement” seeks to enforce compliance with existing laws at grade crossings. The most prominent educational and outreach effort for grade crossing safety in the U.S. is Operation Lifesaver (OL). OL’s network includes certified volunteer speakers and trained instructors offering free rail safety education programs to school groups, driver education classes, community audiences, commercial drivers, law enforcement officers, and emergency responders (Savage, 2006). Mok and Savage (2005) found that the introduction of state-wide Operation Lifesaver programs results in a 15% decrease in grade crossing incidents, and a 19% decrease in fatalities.

Jeng (2005) developed a railroad safety section for inclusion in the New Jersey driver’s manual, then performed an experimental driver’s test on a set of drivers. Drivers who studied the manual with the additional section performed significantly better on the test than those who did not. This suggests that an accurate and easily understood driver’s manual could improve drivers’ behavior at grade crossings.

Research into human factors has shown that engineering solutions solve only part of the problem. Education is also a critical component for reducing collisions at grade crossings. It is important to study the response of drivers to any new type of grade crossing and, if necessary, implement education and awareness initiatives. This is especially the case at shared corridor crossings, where drivers may be accustomed to conventional train speeds and frequencies, but do not expect more frequent, higher-speed passenger trains.

2.6. CONCLUSIONS

Highway-rail grade crossing risk is a topic that has received considerable attention from the highway perspective, but much less from the rail perspective. The results of these studies have led to significant safety advancements, especially in improved train crashworthiness, improved grade crossing design, and expanded driver education. These advancements, combined with increased government oversight, have led to a steady, decades-long reduction in the number of grade crossing incidents and associated casualties in the U.S.

CHAPTER 3: METHODOLOGY AND STATISTICAL BACKGROUND

This chapter discusses the U.S. DOT Federal Railroad Administration (FRA) databases used in this research, and provides a basic background in the statistical techniques and terms used in later chapters.

3.1. PERIOD OF STUDY

Although grade crossing incidents are common, derailments due to grade crossings accidents are relatively infrequent, with an average of fewer than 20 each year. In order to develop a database with a sufficient number of records, I used data for the 20 year period from 1991 to 2010. This provided a robust sample of grade-crossing-caused derailments to enable model development. Data were most recently downloaded from the FRA Safety Data website in 2015. Information in the FRA databases is updated for a period of five years after the end of each calendar year. This suggests that, by 2015, incident data from 2010 should be "stable" and no additional modifications are likely. However, it is possible that future researchers downloading the same years' data will observe small differences in the number of incidents compared to the numbers presented here. This is normal due to how the databases are maintained.

To validate the model, data from 2011 through 2014 were used, which was all the new data available at that point. These data were downloaded in early 2016, and therefore had probably not reached a fully "stable" state. Therefore, if the analysis is repeated in five years, the numbers might differ slightly from those presented here, though the overall trends are expected to be the same.

3.2. FRA DATABASES

The FRA maintains three databases that were used in this study: the Rail Equipment Accident/Incident (REA) database, the Highway Rail Accident (HRA) database, and the Grade Crossing Inventory (GCI) historical file. Since the databases and the fields in them change periodically, the file structures that were current at the time of this study are attached (Appendix H). Future researchers can cross-reference with any updated file structures to find comparable fields.

3.2.1. REA Database

The REA database collects data concerning "events involving the operation of railroad on-track equipment (standing or moving) and causing reportable damages greater than the reporting threshold for the year in which the accident/incident occurred" (FRA, 2011b). This threshold periodically changes to account for inflation and other adjustments; as of 2011 it was set at $9,400. These data are reported to the FRA using the FRA F 6180.54 form, which is filed by railroads that experienced an incident meeting the reporting threshold criterion. It provides useful information about incidents, such as incident cause, number of cars or locomotives derailed, length of consist, type of track involved, and a number of other variables of interest.

The fields primarily used in this study were those concerning the time and location of the incident (IYR, IMO, DAY, TIMEHR, TIMEMIN, AMPM, GXID), incident cause/circumstance (TYPE, CAUSE, TYPEQ), derailed rail vehicles (HEADEND2, MIDMAN2, MIDREM2, RMAN2, RREM2, LOADF2, LOADP2, EMPTYF2, EMPTYP2, CABOOSE2), and train length (HEADEND1, MIDMAN1, MIDREM1, RMAN1, RREM1, LOADF1, LOADP1, EMPTYF1, EMPTYP1, CABOOSE1).

3.2.2. HRA Database

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

The fields primarily used in this study were those concerning the time and location of the incident (IYR, IMO, DAY, TIMEHR, TIMEMIN, AMPM, GXID), highway vehicle characteristics (TYPVEH, VEHSPD), train characteristics (RREQUIP, RRCAR, TYPEQ, TRNSPD), incident type (TYPACC), and warning device (CROSSING).

3.2.3. GCI File

The GCI database includes information reported to the FRA by each state DOT about the condition of each crossing. This includes information about the highway (i.e. annual average daily traffic (AADT), percent truck traffic) and the rail line (i.e. timetable speed, daily number of trains).

There are two types of GCI file available to download from the FRA's data website. The first is a current file, indicating the current conditions at each crossing based on the most recently submitted crossing condition report. The second is a historical file that collects all condition reports. A new crossing condition report is supposed to be submitted each time conditions change at the crossing (such as adding gates or changes in train service), so many crossings have multiple records in the historical file.

In practice, some states have been better about updating the file than others, so data in the GCI may be incomplete or out of date. Information that would come from the railroads (such as number of trains per day) is often particularly sparse. As of 2015, FRA requires railroads to provide information for crossings over which they travel directly to the GCI (FRA, 2017b). This will improve the accuracy of the database for future researchers, though it did not affect the records used in this study.

The fields primarily used in this study were those concerning the location of the grade crossing (CROSSING), the effective dates of the crossing report (EFFDATE, EDATE), number of trains using the crossing (TOTALTRN, PASSCNT), timetable speed (MAXTTSPD), highway characteristics (HWYCLASS, AADT, PCTTRUK), and crossing angle (XANGLE)

As of 2015, the structure of the GCI files has also been changed compared to what was used in this study. Most of the same data are being collected, but the names of some data fields have been changed.

3.3. ADDITIONAL DATA SOURCES

For some analyses, data were needed that could not be obtained from the FRA databases. In these cases, other sources were used, including the Universal Machine Language Equipment Register (UMLER), the Official Railway Equipment Register (ORER), locomotive registers and spotters' guides, railroad photography websites, and Google Earth. These will be explained later in the dissertation.

3.4. COMBINING DATABASES

A contribution of this dissertation is a methodology for combining information from all three databases. Each database contains information that is useful for developing a model of grade crossing derailment likelihood, but none of the databases by itself contains all the information needed. The HRA database includes data about factors that may affect derailment likelihood and severity; however, it does not provide a means of identifying derailment events, which can only be found in the REA database. Additionally, only the GCI provides information about crossing angle and geographical location. Therefore, the first step was to identify a way to combine this information.

The REA and HRA databases have a set of common fields that can be combined to create an identifier. I combined the fields for incident year (IYR), month (IMO), day (DAY), hour (TIMEHR), minute (TIMEMIN) and grade crossing identifier (GXID). This produced a unique code, since it is unlikely that two incidents will occur at the exact same time at a single crossing. This code (IDNO) was used as a relationship key between the two databases in Microsoft Access, the database software used in this analysis. All records in the HRA database were preserved in the analysis. For incidents listed in both the REA and HRA databases, the REA information was added to the HRA record. Arguably the most important fields from the REA database were those indicating the number of rail vehicles derailed. I assumed that if an HRA record had no corresponding REA record, no derailment occurred.

Information from the GCI was incorporated using GXID as the relationship key. The GCI is a historical record, and as such contains many records for each grade crossing. Therefore, when relating information from the HRA database to the GCI, one must select the GCI record that was in effect at the time of the HRA incident. If not, conditions at the crossing could be inaccurate, thereby affecting the results. I selected the GCI record whose effective (EFFDATE) and end (ENDDATE) dates bracket the incident date from the HRA[1]. Due to incomplete reporting, many HRA records did not have corresponding information in the GCI. For example, of the approximately 44,000 freight train grade crossing collisions in the HRA database, only 2,221 had complete records in the GCI. Analysis that relied on GCI information was conducted using only complete records.

The complete GCI data might not be representative of all grade crossings. There may be reasons that inventory data were collected for these crossings but not for others. For example, more complete data may be provided by some states than others. Additionally, some road types, such as major arterials, may be evaluated more frequently due to their importance, resulting in better data quality for these roads compared to local roads. If this is the case then it would contribute unexplained variance to the results.

It was also important to determine which records were internally inconsistent and to eliminate records that could not be used in the analysis. Records that were missing values for the most important factors (incident type, vehicle speed, vehicle type, and train speed) were excluded. Additionally, records with inconsistent values were excluded. For example, if an incident was reported as “vehicle strikes train” and the vehicle speed was reported as zero, then the incident was excluded since such an incident is impossible.

[1] These two fields (EFFDATE and ENDDATE) were removed from the GCI when the FRA updated its data collection methodology in 2015. However, the fields can be recreated using the new field “RevisionDate” by assuming that RevisionDate is equal to EFFDATE for the record, and the day before RevisionDate is equal to ENDDATE for the previous record at a crossing.

3.5. STATISTICAL BACKGROUND

3.5.1. The SAS LOGISTIC Procedure

The models in Chapters 5 and 6 were developed using the LOGISTIC procedure in the Statistical Analysis Software (SAS) program. This procedure uses the method of maximum likelihood to fit a linear logistic regression model to binary response data (SAS Institute, 2013). In this way, the relationship between certain explanatory variables and the outcome responses can be analyzed. In the case of grade crossing incidents, for each incident record the output of the model is a value between 0 and 1 representing the probability of a derailment occurring. Logistic regression is generally discussed in terms of "events" and "non-events"; in the analysis described here, a derailment is an event, and an incident in which no derailment occurs is a non-event.

The SAS LOGISTIC procedure has five effect selection techniques: none, forward selection, backward elimination, stepwise selection and best subset selection (SAS Institute, 2013). "None" fits the full specified model, so if four variables are provided, all four will remain in the model. Forward selection adds each statistically significant variable to the model in sequence according to the strength of its influence. Backward elimination begins with the full specified model, then removes non-significant variables in order from least significant. In stepwise selection, the variable having the strongest influence on the response variable is added at each step. Variables are not added if they are found to have an insignificant influence on the model. At each step, the procedure tests the influence of including each factor by performing a "goodness-of-fit" test. It also examines the factors that have already been added and removes any that are found to no longer have a significant effect. The best subset option identifies a number of models with the highest likelihood scores out of all potential models. This method does not work with categorical variables and so was not used. Stepwise selection was used to develop the models in Chapters 5 and 6, though it was found to produce the same models as forward selection and backward elimination. SAS indicates any co-linear variables to the user, as highly correlated variables can impair the ability of SAS to identify the best variable for inclusion in the model.

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

Additionally, SAS LOGISTIC has an option called SCORE that can be used when validating a logistic regression model. SCORE allows the user to specify a set of unseen data that the software analyzes using the model developed by the training data.

3.5.2. Evaluating the Fit of a Logistic Regression

Once a model has been created using logistic regression, there are a variety of metrics that can be used to assess how well the model “fits” the data. The SAS LOGISTIC procedure produces several of these metrics. The following paragraphs describe the tests used to evaluate fit: the Hosmer-Lemeshow (HL) goodness-of-fit test (for fitted data only), the area under the Receiver Operating Characteristics (ROC) curve, and the Brier test (for scored data only)

3.5.2.1. Hosmer-Lemeshow (HL) Goodness-of-Fit Test

The Hosmer-Lemeshow goodness-of-fit test returns a value between 0 and 1 indicating how well the model performed at predicting the number of events in various random subsets of the input data set. Values closer to 1 indicate good model fit, and values closer to 0 indicate poor fit (Hosmer and Lemeshow, 2002). A value of 0.5 essentially indicates that the model performs no better than flipping a coin or random guessing.

3.5.2.2. Area Under the Receiver Operating Characteristics Curve (AUC)

The Area Under the ROC curve (AUC) is a value between 0 and 1 that describes the ability of the logistic regression model to discriminate between events and non-events for any chosen threshold. The threshold is a value selected by the model's user to represent the distinction between an event and a non-event. Since the logistic model returns a value anywhere between 0 and 1, it is a matter of statistical or engineering judgment to choose the threshold, though the "best" threshold will generally be equal to the proportion of events in the dataset. Any datum with a value less than the threshold will be considered a non-event, and any datum with a value greater than the threshold will be considered an event. The AUC value is important because it quantifies the model's ability to distinguish between events and non-events regardless of the chosen threshold. The ROC curve plots sensitivity versus (1 – specificity) for every threshold value. Sensitivity can be thought of as the "true positive rate", or a measurement of actual events divided by the number of events identified by the model. Analogously, specificity is the "true negative rate", or a measurement of the number of actual non-events divided by the number of non-events identified by the model.

As with the HL test, an AUC closer to 1 indicates a model that better discriminates between events and non-events; if the AUC is 0.5, then it is no more accurate than random chance. The positive diagonal 1:1 line shown on the ROC curves in Chapters 5 and 6 indicates what the ROC curve would look like if the AUC was 0.5. Generally, a model is considered to provide acceptable discrimination if the AUC value is between 0.7 and 0.8, good discrimination for values between 0.8 and 0.9, and excellent discrimination if the value exceeds 0.9 (Hosmer and Lemeshow, 2002).

3.5.2.3. Brier Score

The SAS software does not provide an HL value for scored data; however, it does provide a Brier score. The Brier score of a data set is a measure of goodness-of-fit; it represents the difference between the predicted probability and the observed response of a data point. Brier scores range from 0 to 1, with the best possible value being 0.

3.5.3. Rare Events Logistic Regression (RELR)

When logistic regression is used on data that have many more non-events than events, the regression will produce a poor fit even though there are indications of strong statistical relationships in the data. The models predict non-events correctly at the expense of predicting events, since this reduces the error rate. In this way, the model predicts a large percentage of all events correctly, but has poor fit because it fails to predict most derailment events.

This problem can be addressed by using a modified form known as “rare events logistic regression” (RELR) (King and Zeng, 2001; van den Eeckhaut et al., 2006). RELR corrects for the disproportionate number of non-events by selecting a random subset of them equal to 1 to 5 times the number of events. In this research, a dataset was created containing a number of randomly-selected, non-derailment events equal to twice the number of derailment events.

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

CHAPTER 4: DESCRIPTIVE STATISTICS OF HIGHWAY-RAIL GRADE CROSSING INCIDENTS

4.1. INTRODUCTION

This chapter presents statistical analysis examining the basic characteristics of grade crossing incidents. Three databases maintained by the Federal Railroad Administration (FRA) were used to understand the effect that grade crossing collisions have on trains, especially as they affect train derailments. Although some highway-rail grade crossing collisions result in derailment of the train, most do not; the challenge is to identify the critical factors affecting the former. This chapter begins by describing some of the characteristics of grade crossing-caused derailments, comparing them to other railroad incident types, and seeks to answer the following questions:

- How severe are grade crossing collision-caused derailments, and how do they compare to other derailment causes?
- Does vehicle size affect derailment likelihood?
- Does vehicle size affect derailment severity?
- Does impact velocity (of the highway vehicle or train) affect derailment rate or severity?
- How does vehicle size affect the speed distribution of a vehicle striking a train?
- Does the weight of the rail vehicle involved in the collision affect derailment likelihood?
- Does the collision angle of the train and highway vehicle affect derailment likelihood?

4.2. METHODOLOGY

The analysis in this chapter uses data from the Highway-Rail Accident (HRA), Railroad Equipment Accident (REA) and Grade Crossing Inventory (GCI) databases collected from 1991 through 2010 (described in Chapter 3). The HRA database contains the largest amount of pertinent information; however, it does not provide information about the number of cars or locomotives derailed in the incident. On the other hand, the REA database provides derailment information but lacks detailed data regarding grade crossing incidents. The GCI provides additional information about crossing characteristics that is useful for the crossing angle analysis.

The three databases were merged to create a unique dataset consisting of incidents that were reported using both the HRA and REA forms. A unique identification code was created for each incident in the HRA and REA databases. The code concatenates the date, time, and crossing identification number (GXID) for each incident to provide a field that can be cross-referenced between the two databases. The GCI information is cross-referenced by looking at the GXID and effective date range of each record.

This methodology results in a consolidated dataset consisting only of incidents that occurred at grade crossings and were also REA-reportable (i.e. exceeded the REA damage value threshold). This consolidated dataset contains what are likely the most severe grade crossing incidents. Mainline grade crossing incidents are the focus of this study because they account for approximately 88% of all incidents, and because mainline derailments are the ones most likely to result in major consequences.

4.3. RESULTS

From 1991 through 2010, 71,153 total grade crossing incidents occurred in the U.S. Of these, 59,893 occurred at public grade crossings on mainline tracks, including 3,135 incidents that exceeded the FRA reporting threshold for track and equipment damage and therefore were additionally reported to the REA dataset. These 3,135 incidents occurred at 2,774 distinct grade crossings. About 90% of crossings had only one incident during the study period, 8% had two incidents, and 2% of crossings had three or more incidents. Of the REA-reportable incidents, 399 resulted in derailment. Over the study period, mainline grade crossing derailments resulted in:

- 399 derailments
- 2,185 freight and 235 passenger cars/locomotives derailed
- 138 railcars carrying hazardous materials damaged or derailed
- Nine hazardous materials railcars released
- 29 highway truck hazardous materials spills (such spills can also be caused by grade crossing incidents that do not result in derailment)
- Evacuation of over 1,000 people due to safety precautions related to hazardous materials spills

The frequency distribution of incident severity as measured by the number of cars or locomotives derailed in individual grade crossing incidents was plotted (Figure 4.1). The modal value was for incidents in which one car or locomotive derailed with declining frequency up to a maximum of 36 derailed cars and locomotives in one incident.

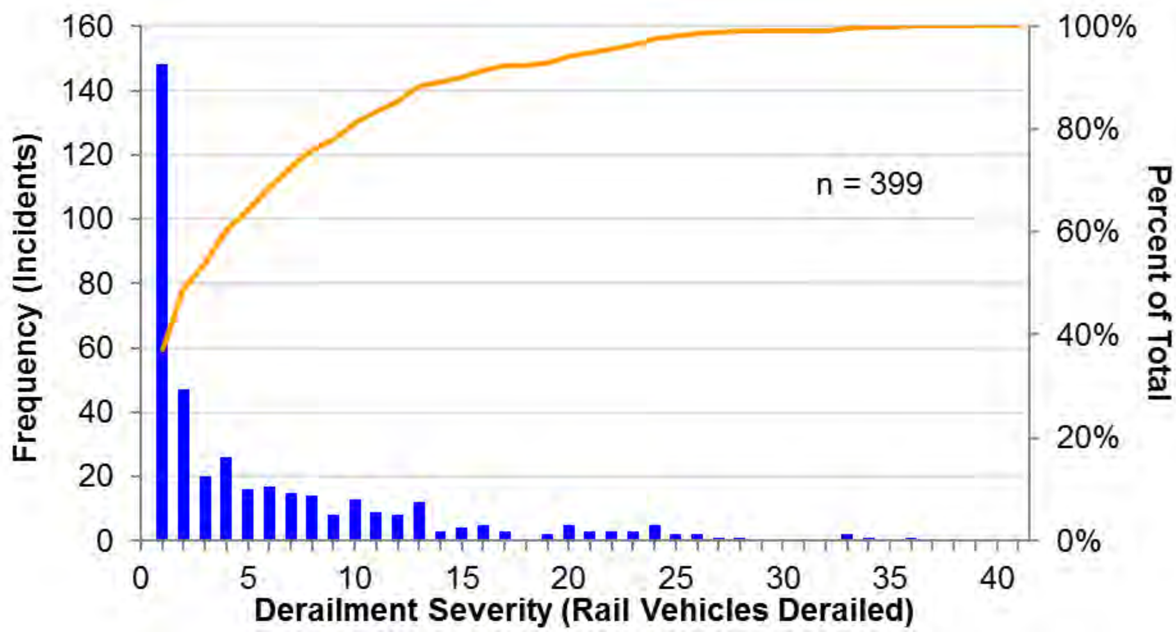


Figure 4.1: Number of rail vehicles derailed per incident at grade crossings given that a derailment occurred (71,153 grade crossing incidents occurred in total).

4.3.1. Grade Crossing Incident Frequency and Severity

When discussing grade crossing incidents in the context of railroad safety, it is useful to compare the frequency and severity of such incidents to other railroad incident causes (Barkan et al., 2003; Liu et al., 2012). The frequency and average number of rail vehicles derailed was calculated for each incident cause (Figure 4.2). Most REA-reportable incidents are derailments (FRA 2011a). Grade crossing incidents are an exception, in that most do not result in derailments. Therefore, two points are shown for grade crossing incidents: one for derailments only (closed circle) and one for all grade crossing collisions in the REA database (open circle).

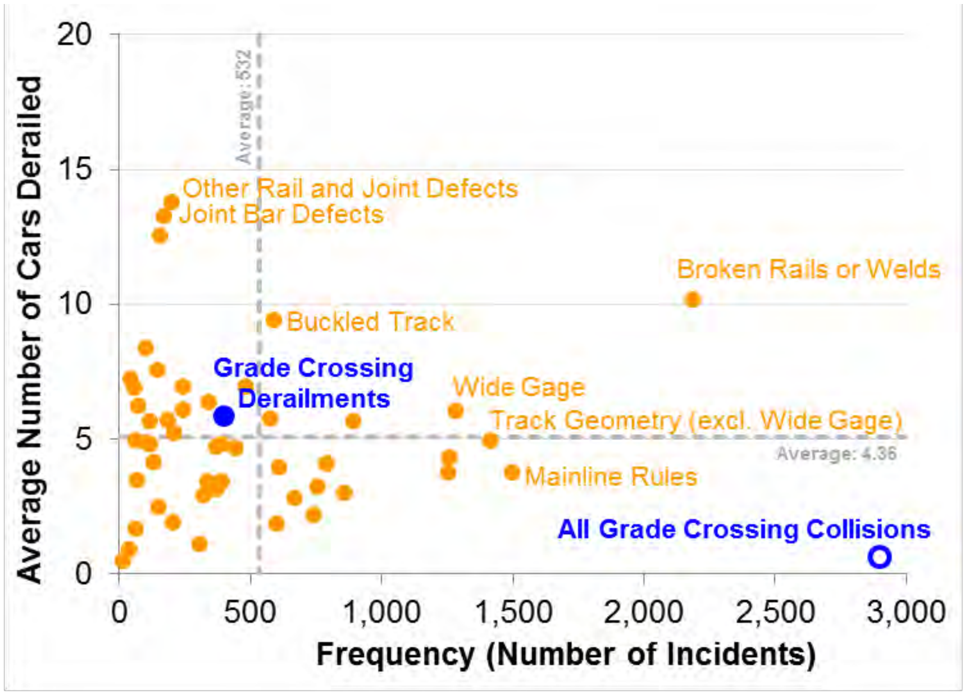


Figure 4.2: Railroad Incidents by Cause Severity vs. Frequency, 1991-2010 (Note: figure uses data from the REA database, therefore only REA-reportable grade crossing incidents are represented).

Comparing other railroad incident causes to all REA-reportable grade crossing incidents shows that they are the most common; however they have a low derailment rate (only 399 of the approximately 2,900 REA-reportable grade crossing incidents resulted in derailment) and consequently low severity. However, if the comparison is made to only those grade crossing incidents that resulted in derailment, they are less frequent than many other incident causes but have above average severity.

Combined, this means that grade crossing collisions having the potential for catastrophic consequences including passenger fatalities and the release of hazardous materials, if a derailment does occur. Even if this does not happen, the frequency of grade crossing incidents results in large societal costs due to property damage and loss of human life (they are the second largest cause of railroad fatalities, after trespassing), as well as significant lost time to railroad operators.

4.3.2. Incident Type

The FRA database differentiates between grade crossing incidents in which the train struck the highway vehicle (TSV) and the highway vehicle struck the train (VST)[1]. As discussed in Chapter 1, incident type is an important factor in my analysis as the factors affecting derailment likelihood are different for the two types.

Derailment likelihood also varies with respect to incident type. VST incidents represent 22% of all grade crossing incidents, but 32% of grade crossing-caused derailments (Figure 4.3). This means VST incidents are disproportionately more likely to result in derailments (χ^2 test: $p < 0.0001$).

[1] This is not a standard terminology, and others refer to these by different names. In the HRA database, these are referred to as "rail equipment struck highway user" (TYPACC = 1) and "rail equipment struck by highway user" (TYPACC = 2) (FRA, 2011b).

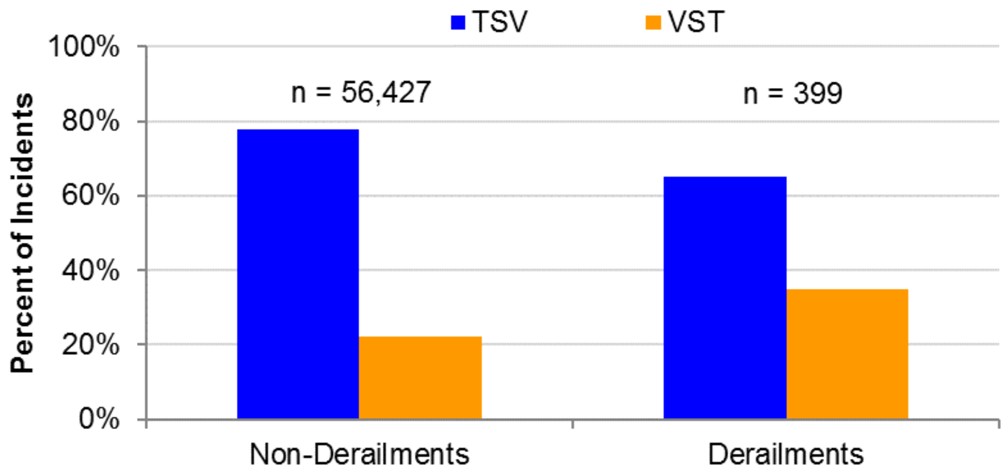
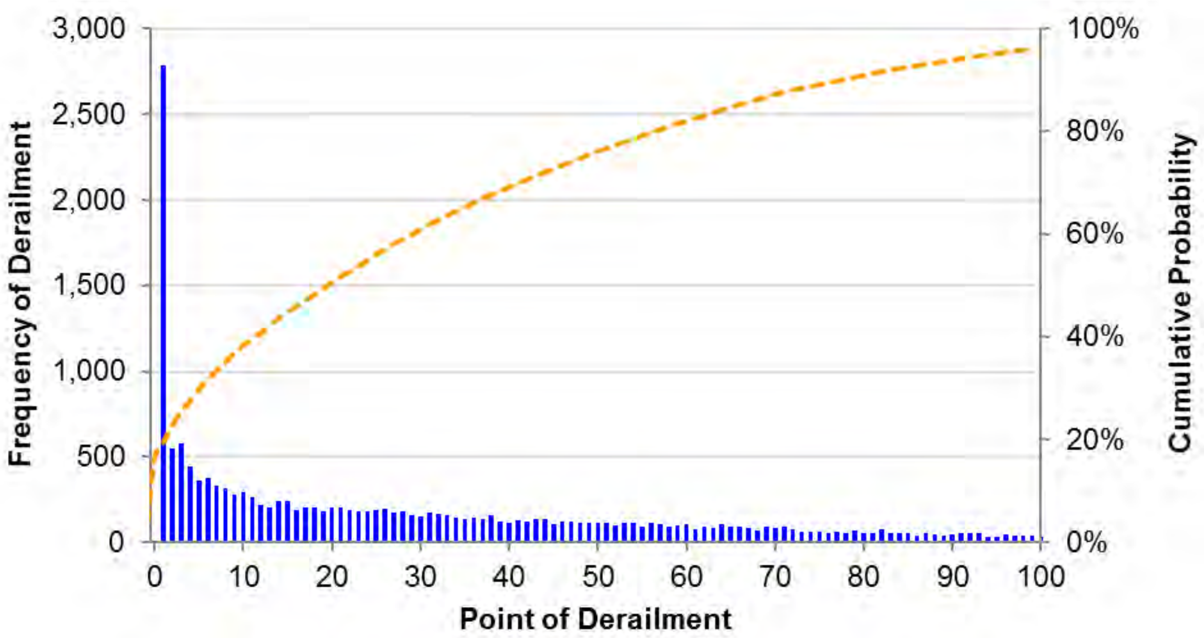


Figure 4.3: Incidents occurring at grade crossings on mainline track from 1991 to 2010 by incident type (TSV or VST).

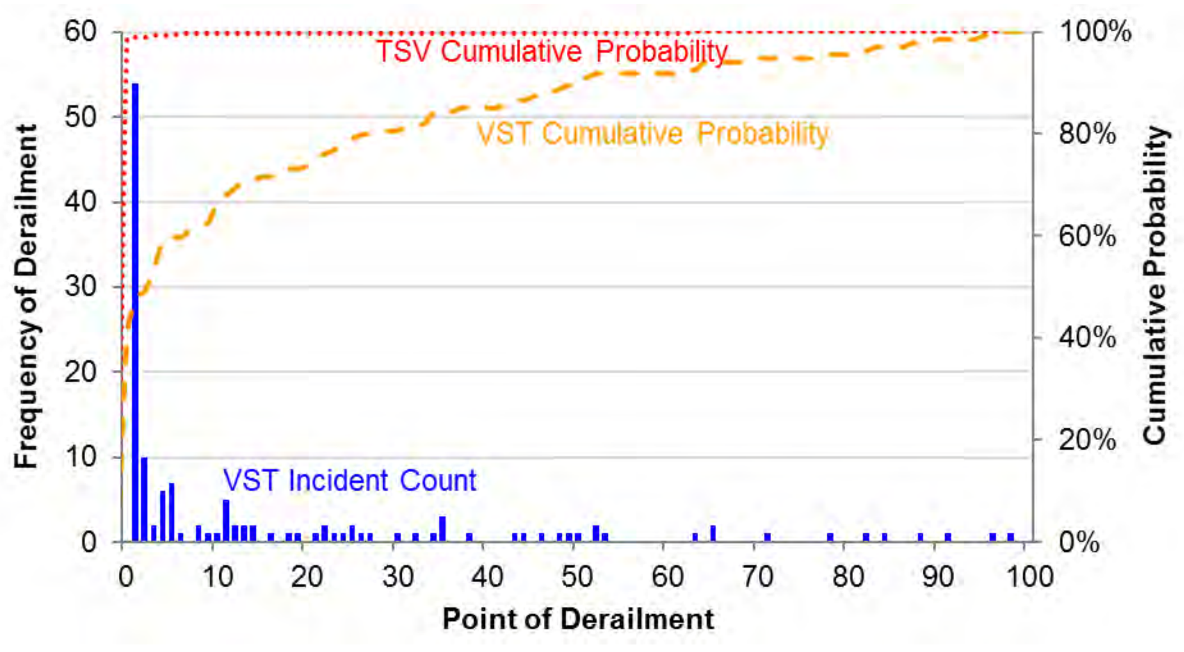
4.3.3. Comparison of Point of Derailment against Various Incident Types

To better understand how grade crossing-caused derailments compare to other derailment causes, an analysis was performed examining the point of derailment (POD) distributions for grade crossing and other incidents. The POD is the position in the train of the first (closest to the front of the train) rail vehicle to derail in an incident.

Figure 4.4a shows, for all incident types, the frequency and cumulative distribution of the position-in-train of the first rail vehicle derailed. Approximately 20% of incidents involve derailment of the front-most locomotive or car. In contrast, Figure 4.4b shows the same chart but for grade crossing incidents alone. In these incidents, most – nearly 80% – involve the front-most locomotive or car in the train. Therefore, grade crossing incidents are more biased to the front of the train compared to all incident causes. Additionally, the POD distribution is different for VST and TSV incidents; while TSV incidents almost exclusively involve the first few rail vehicles in the train, VST incidents are more distributed along its length.



(a)



(b)

Figure 4.4: Frequency and cumulative distribution of the POD for (a) all derailment causes, and (b) grade crossing caused incidents (by incident type, TSV or VST), 1991-2010.

Train lengths vary widely depending on train type (passenger or freight) and other factors. As a result, shorter trains will naturally have a POD closer to 1, which exaggerates the front-of-train bias. To account for this, the normalized point of derailment (NPOD), which is the position-in-train of the first derailed car divided by the train length, can be used. This distribution was calculated for a variety of incident cause groups to show the effect of incident cause on NPOD (Figure 4.5).

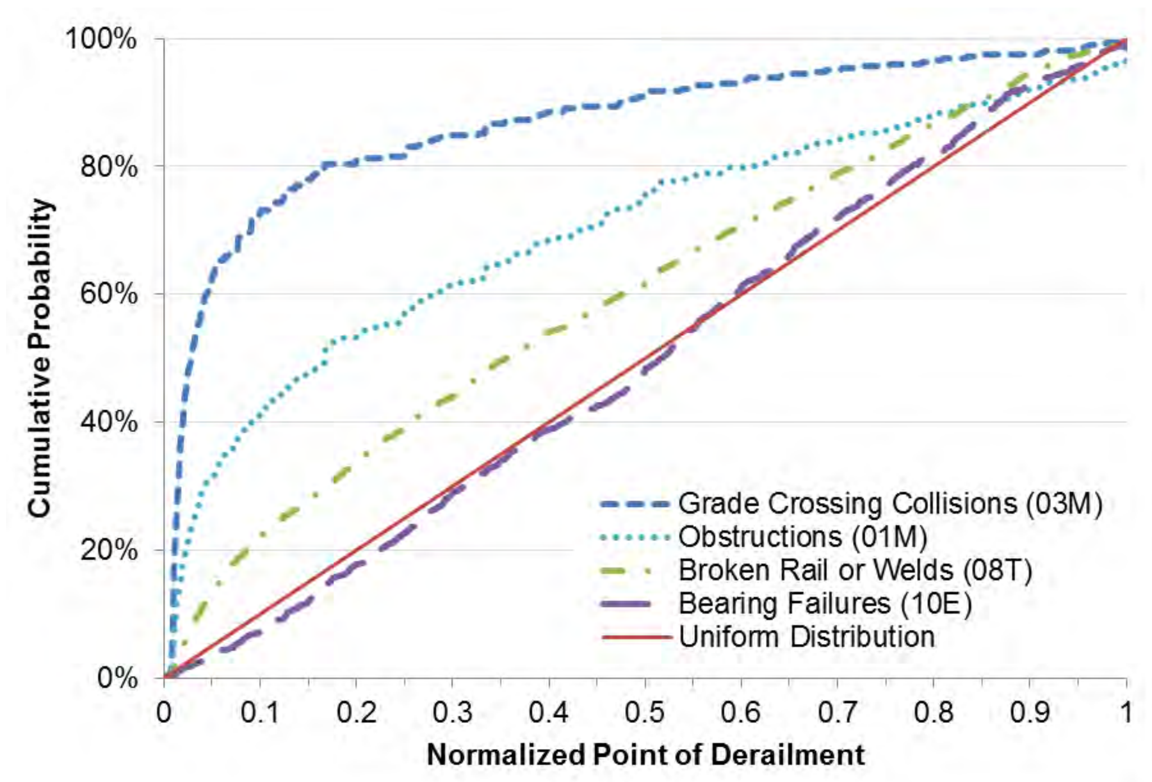
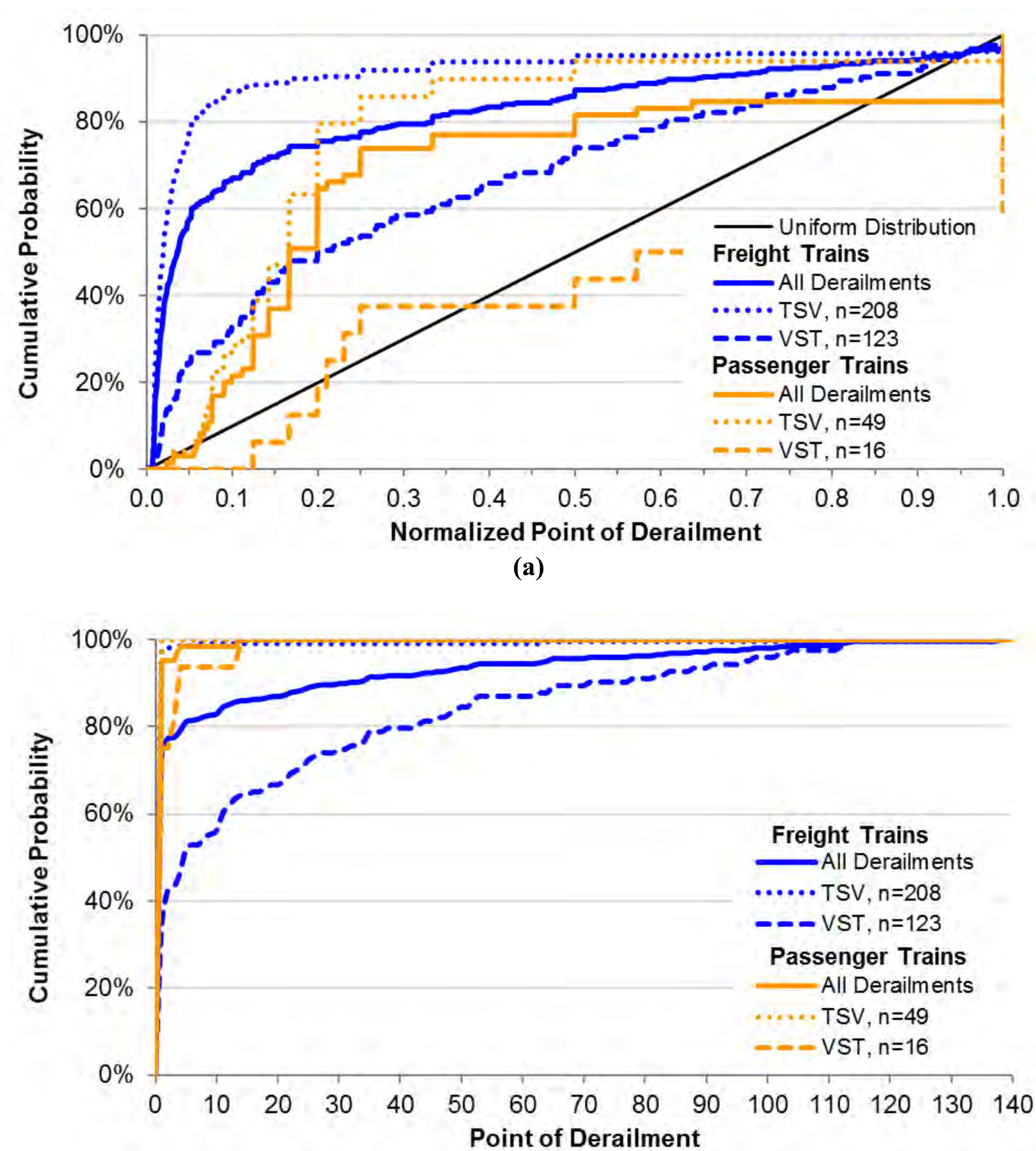


Figure 4.5: Cumulative Distribution Probabilities for normalized POD of various accident causes.

Of the four incident cause groups examined, grade crossing collisions were the most heavily biased towards front-of-train derailments. In comparison, incidents with the cause group “Obstructions” are moderately biased towards the front of the train, and “Broken Rail or Welds” are also somewhat biased towards the front. In contrast, derailments due to "Bearing Failures" are uniformly distributed throughout the train. To illustrate the difference in NPOD distributions, consider that 75% of grade crossing collisions have an NPOD less than or equal to 0.1, whereas the same is true for only 43% of obstructions, 23% of broken rails, and 8% of bearing failures. If NPOD was uniformly distributed, we would expect 10% of incidents to have NPOD less than or equal to 0.1. It is evident that derailments resulting from grade crossing incidents are more likely to occur at the front of the train compared to other causes.

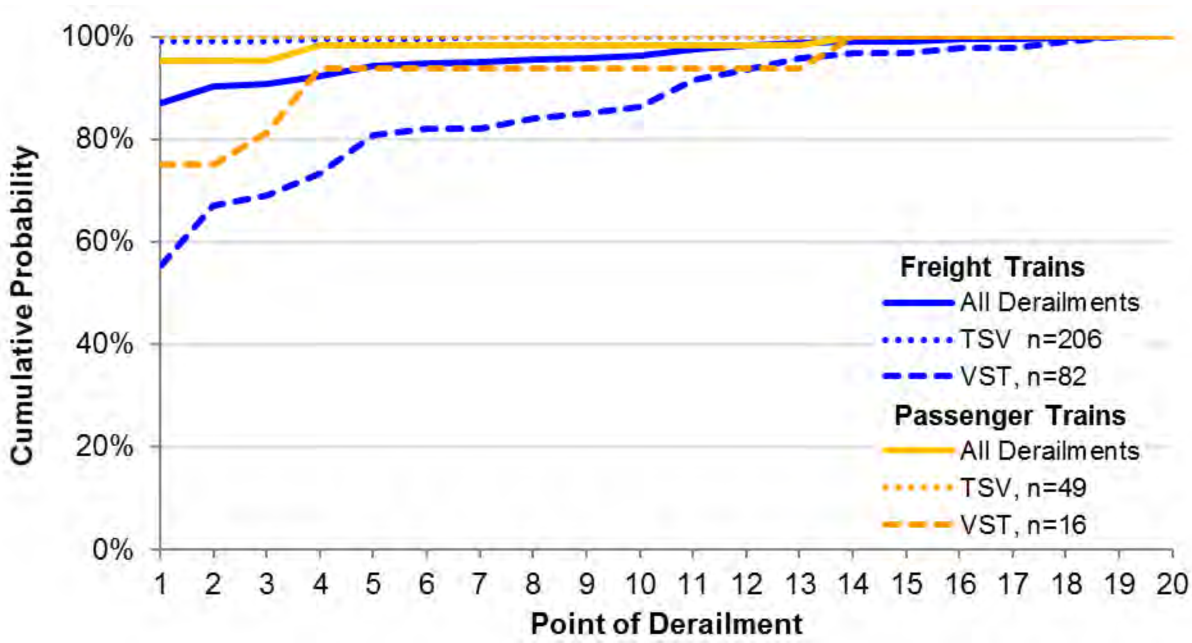
The NPOD analysis was repeated for grade crossing incidents alone (Figure 4.6a). The data were split into categories according to freight- and passenger-involved derailments, and further divided into TSV and VST incidents. Freight train-involved incidents might appear to be more strongly biased toward the front of the train; however, since this is a normalized point of derailment, this difference is mostly due to the fact that passenger trains are much shorter than freight trains. Given a 50-car freight train and an 8-car passenger train, if the first car in each were to derail, the freight train would have a normalized POD of 0.02 and the passenger train would have a normalized POD of 0.125. This is illustrated by plotting the non-normalized POD for freight and passenger trains (Figure 4.6b). No passenger train derailments affected a point in the train farther back than the 20th rail vehicle, and only 13% of freight train derailments had a POD greater than 20. Therefore, I also plotted the POD distributions considering only incidents with a POD less than 20 (Figure 4.6c). It is evident that POD trends are similar for freight and passenger trains.



(a)

(b)

Figure 4.6: Cumulative Distribution Probabilities for (a) normalized POD of grade crossing collisions (03M) by equipment and incident type, (b) non-normalized POD for all train lengths.



(c)

Figure 4.6 (cont.): Cumulative Distribution Probabilities for (c) non-normalized POD for trains shorter than 20 rail vehicles[2].

Of greater interest is the effect of incident type on POD distribution. For both passenger and freight trains, derailments were more biased towards the front of the train for TSV incidents compared to VST incidents. This is not surprising, given that if a train strikes a vehicle the rail vehicle in the leading position is most likely to derail. If the train is struck by a highway vehicle, it will affect whatever part of the train is struck. Consequently, there is lower likelihood of the POD being near the front of the train compared to TSV incidents. However, even VST incidents still show some bias towards the front of the train.

---

[2] Three derailments involved “maintenance-of-way equipment” (TYPEQ=A) and were excluded from this analysis.

4.3.4. Large Highway Vehicle Involvement

An underlying physical factor discussed in Chapter 1 was the mass of the highway vehicle. The weight of the highway vehicle is not reported in the HRA database, but there is a field (TYPVEH) indicating the type of highway vehicle that provides some insight (FRA, 2011a, 2011b). This field defines 11 categories of highway user, including automobiles, semi-tractor-trailers, buses, motorcycles and pedestrians. For the purposes of this study, incidents involving straight trucks and tractor-semitrailers (categories B and C) – believed to be the two heaviest categories of highway vehicle – were defined as "large highway vehicles" and all others were defined as "small highway vehicles" [3]. Incidents were omitted if they were classified as "other motor vehicles," "pedestrian" or "other", as is further explained in Appendix A. Removing these categories reduced the size of the dataset by about 2,500 entries. It is possible that about half of these could be added back to the dataset if the narrative fields were used to code the vehicle type manually; however, the dataset is large enough that these entries are unnecessary.

The larger mass of trucks suggests that train collisions in which they are involved may be more likely to result in a derailment (Chapter 1). An analysis was conducted to test this hypothesis and quantify the relative difference between large and small highway vehicles. The total number of REA-reportable, mainline grade crossing derailments involving large vehicles were compared with those involving other vehicles (Figure 4.7).

---

[3] It is possible that buses (type "F") should also be included in the large vehicle category; however, analysis showed that their exclusion did not affect the results for the 20-year period being studied because no derailments exceeding the REA reporting threshold involving buses occurred.

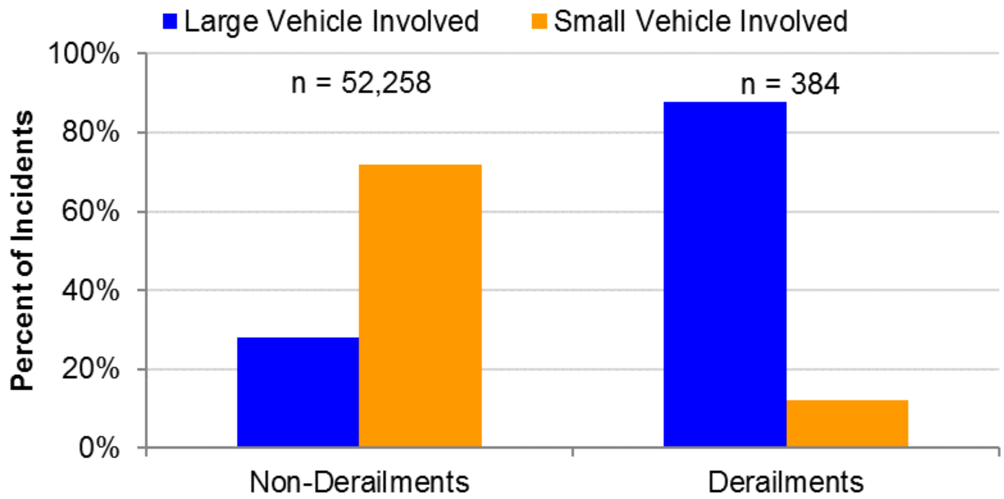


Figure 4.7: Incidents occurring at grade crossings on mainline track from 1991 to 2010 involving large highway vehicles versus all other vehicles[4].

Large highway vehicles were involved in 31% of all mainline grade crossing incidents and 91% of mainline REA-reportable grade crossing derailments and thus were three times more likely to cause a derailment (χ^2 test $p < 0.0001$). The other 9% of grade crossing derailments involved automobiles, pick-up trucks, other motor vehicles, and vans.

The greater tendency for large vehicles to cause derailments led to the question about whether they might also affect derailment severity. To investigate this hypothesis, the distribution of total cars and locomotives derailed in incidents was compared for incidents that did and did not involve large vehicles (Figure 4.8). There was no significant difference between the severity of derailment incidents involving large vehicles and the severity of derailment incidents not involving large vehicles (Wilcoxon Rank Sum (WRS) Pr<Z = 0.0753; t-test p =

[4] Fifteen derailments had a reported highway vehicle type of H (motorcycle), J (other motor vehicle) or M (other) and were therefore excluded from this analysis.

0.1775). In other words, once a motor vehicle has caused a derailment, the severity of that derailment is little affected by the size of the vehicle that caused it.

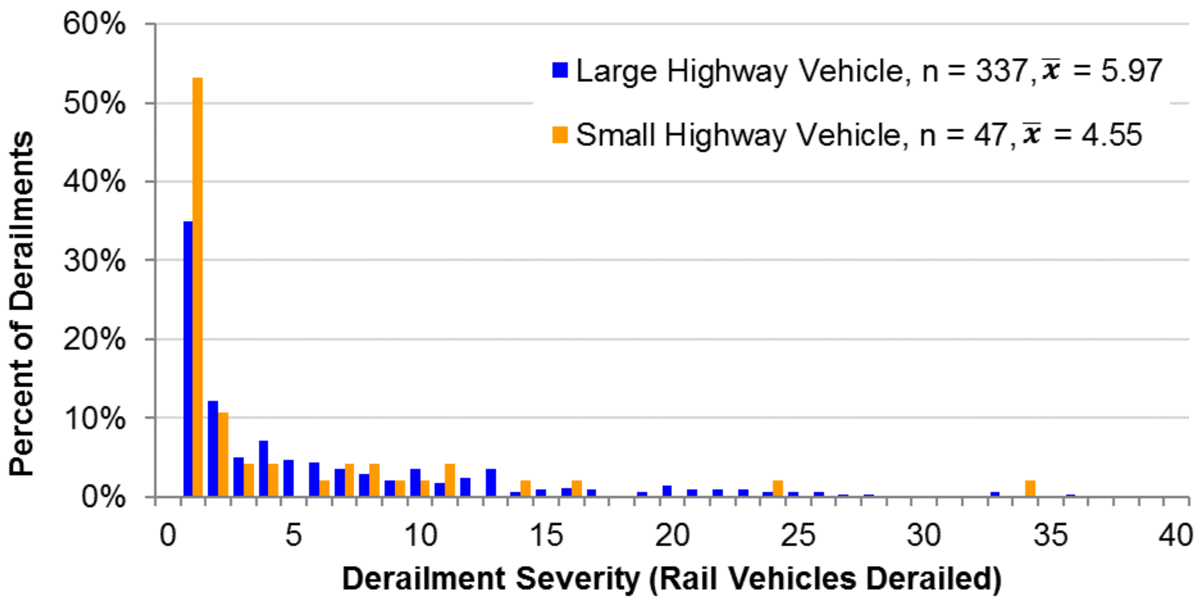
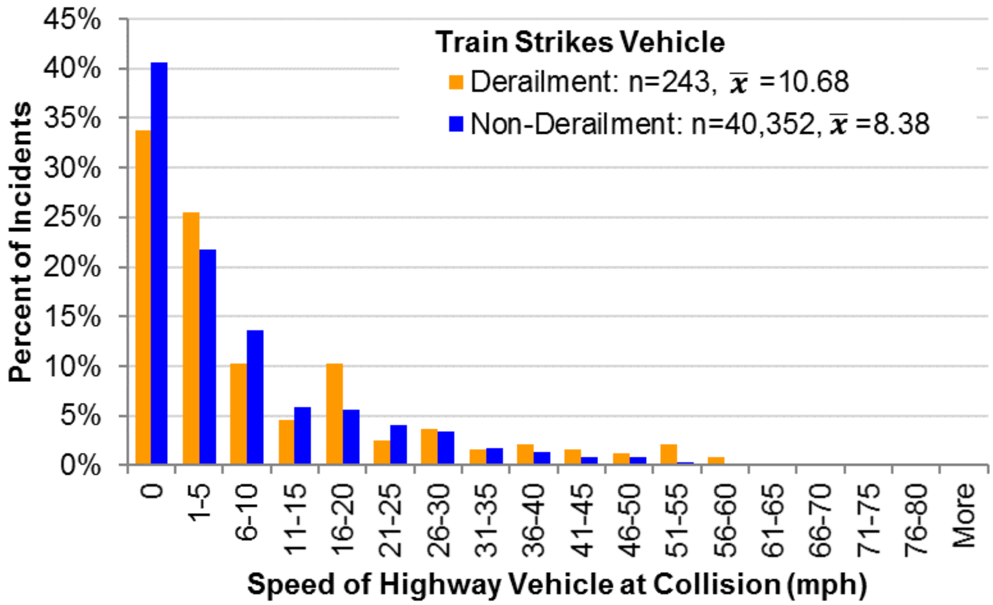


Figure 4.8: Number of cars and locomotives derailed in grade crossing incidents by highway vehicle type. Frequencies are given as a percentage of all incidents of each type.

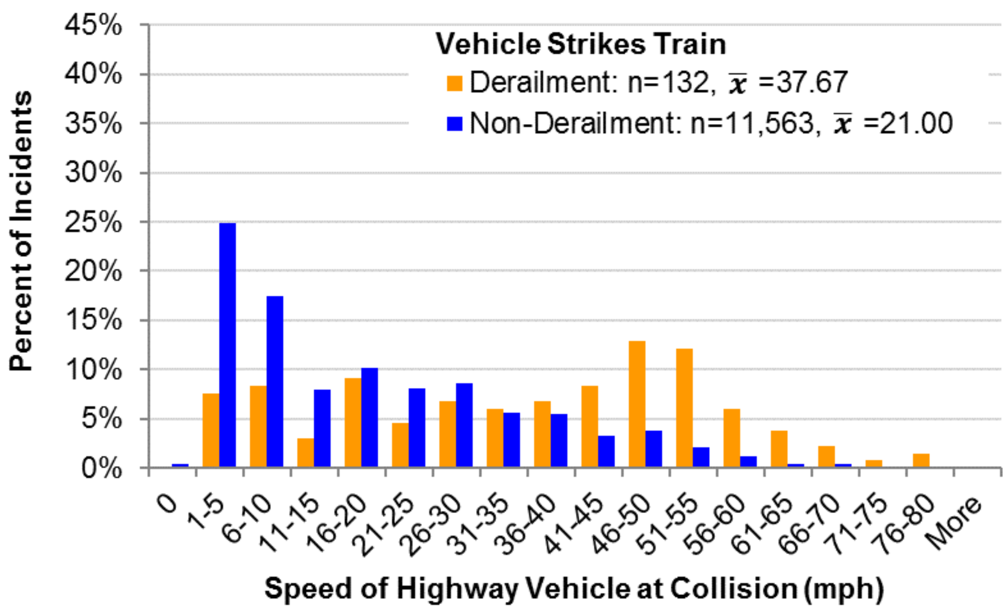
4.3.5. Speed of Highway Vehicle at Collision

Another physical factor of interest is highway vehicle speed. The FRA records data about the speed at collision of the highway vehicle involved in grade crossing incidents. These data are typically estimated by observers at the incident scene and thus are subject to some uncertainty.

The data were divided in two categories based on incident type (TSV or VST). As explained in Chapter 1, the physical mechanism involved in these two types of collisions is likely very different and will be accounted for differently in the final statistical model. For each incident type category, the data were further divided into derailments and non-derailments. Pair-wise comparisons were performed within the incident type categories (Figures 4.9a and b).



(a)



(b)

Figure 4.9: Speed at collision of vehicles involved in grade crossing incidents, 1991-2010 for (a) train striking highway vehicle (TSV) scenario and (b) highway vehicle striking train (VST) scenario[5].

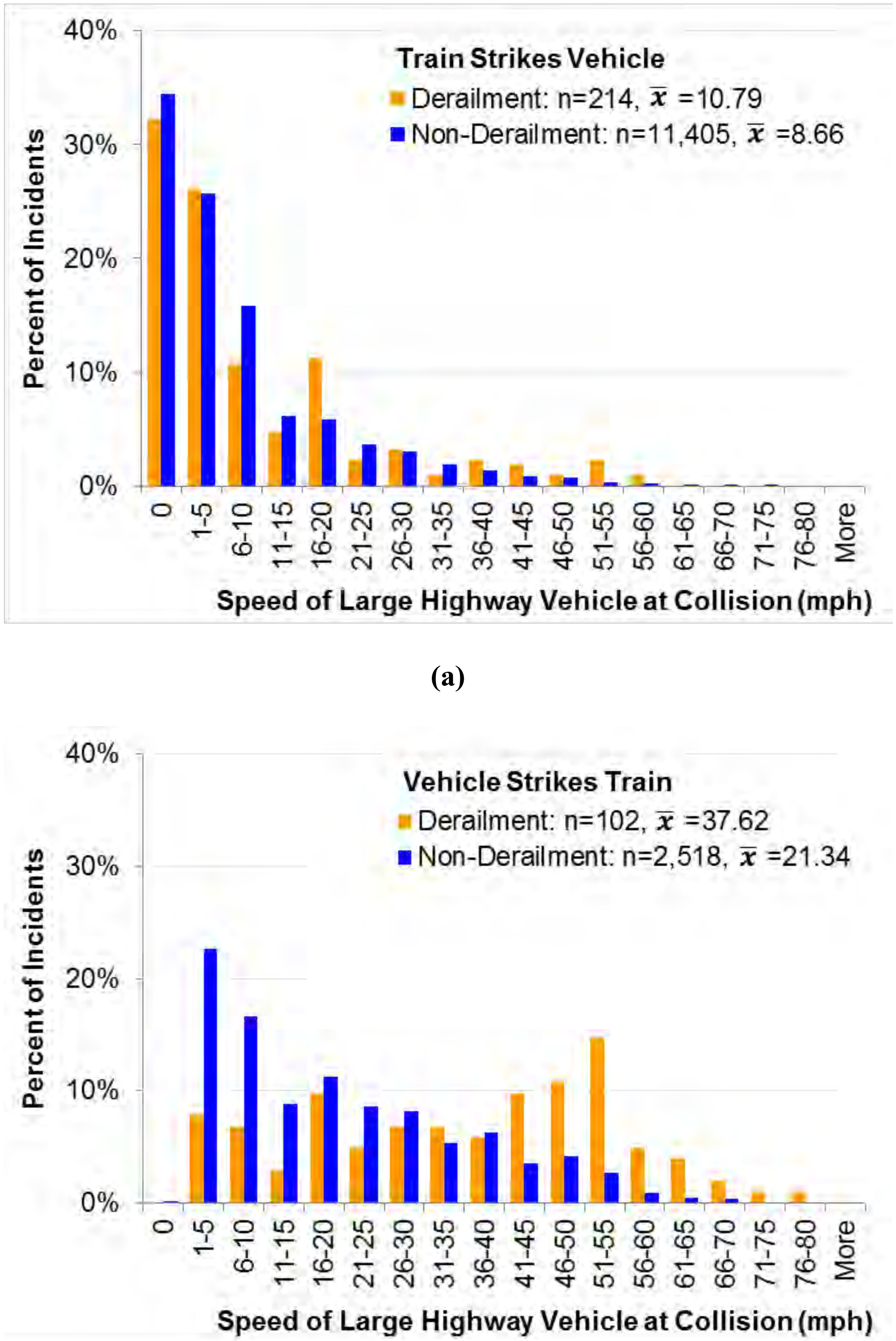
[5] Twenty-four derailment incidents had no reported highway vehicle speed and were not included in this analysis.

The majority of TSV incidents occurred at highway vehicle speeds less than 10 mph. A large number (37%) of incidents occurred in which the highway vehicle was stopped on the tracks. Speeds for the VST incidents were generally higher, with many incidents in the 40 to 60 mph speed range.

The data were analyzed to determine if there was a difference in speed between derailment and non-derailment incidents. For both the TSV (WRS test one-sided $Pr < Z = 0.0135$) and VST (WRS test one-sided $Pr < Z < 0.0001$) derailments were more likely to occur at higher vehicle speeds, but the trend is stronger for VST incidents (Figure 4.9b).

The effect of vehicle speed on derailment severity was also examined. The distributions of the total number of cars and locomotives derailed do not suggest a strong relationship between highway vehicle speed at collision and derailment severity (small highway vehicles – TSV: WRS $Pr > F = 0.1786$, VST: WRS $Pr > F = 0.1506$; large highway vehicles – TSV: WRS $Pr > F = 0.5543$, VST: WRS $Pr > F = 0.0663$; $\alpha = 0.05$).

As discussed in Section 4.3.4, large highway vehicles are involved in a disproportionately large percentage of grade crossing derailments, so the highway vehicle speed analysis was repeated using only large highway vehicles (Figures 4.10a and b). The results for the VST data were the same as in the all-vehicle analysis (WRS test one-sided $Pr < Z < 0.0001$). However, for the TSV data, vehicle speed had no effect on derailment likelihood (WRS test one-sided $Pr < Z = 0.1256$).



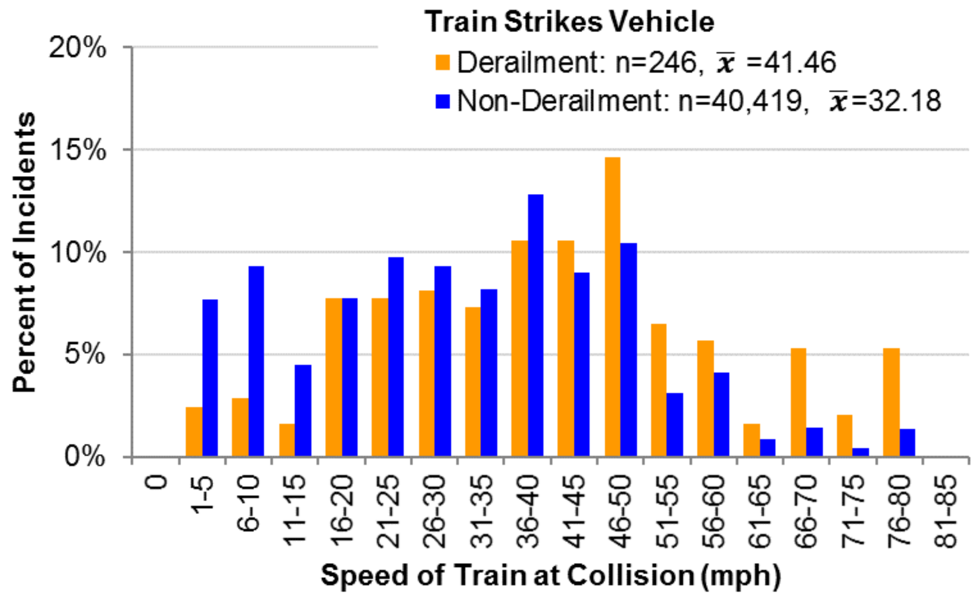
(a)

(b)

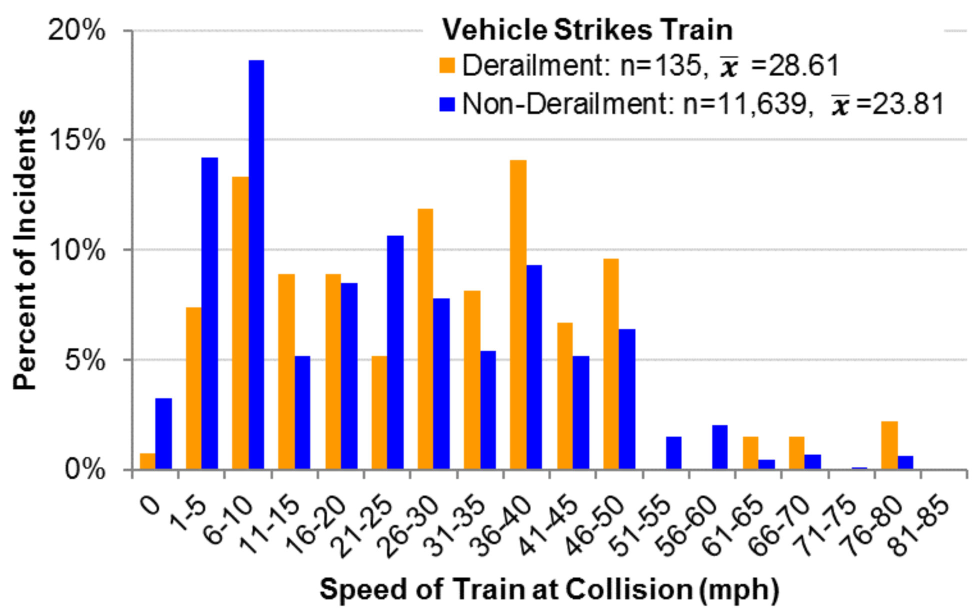
Figure 4.10: Speed at collision of large vehicles involved in grade crossing incidents, 1991-2010 for (a) train striking highway vehicle scenario and (b) highway vehicle striking train scenario.

4.3.6. Speed of Train at Collision

Train speed is also of interest. An analysis was conducted to investigate the effect of train speed on derailment occurrence and severity. The FRA records information about the speed of trains involved in grade crossing incidents, and these data may be either exact or estimated. The data were grouped into the same four categories described in Section 4.3.5, and the percentage of each type of incident that occurred at a given train speed was plotted (Figure 4.11).



(a)



(b)

Figure 4.11: Speed at collision of trains involved in grade crossing incidents for (a) train striking vehicle and (b) vehicle striking train scenarios[6].

[6] Eighteen derailment incidents had no reported train speed and were not included in this analysis.

The distributions for all four scenarios were roughly the same, with the majority of collisions occurring in the 35 to 55 mph range. This is probably because this represents the typical range of mainline train speeds. For both the TSV (WRS test one-sided $Pr < Z < 0.0001$) and VST (WRS test one-sided $Pr < Z = 0.0005$) scenarios, derailments were more likely to occur at higher train speeds.

Since derailment severity is known to increase with train speed for other incident causes, a linear regression was performed to see if the same would be observed for grade-crossing-caused derailments. No relationship between train speed and derailment severity was observed for incidents involving small highway vehicles (TSV: $Pr > F = 0.4456$, VST: $Pr > F = 0.3449$; $\alpha = 0.05$); however, a relationship was observed for incidents involving large highway vehicles (TSV: $Pr > F = 0.0006$, VST: $Pr > F = 0.0015$; $\alpha = 0.05$). Specifically, the number of rail vehicles derailed increases with increased train speed (Figure 4.12).

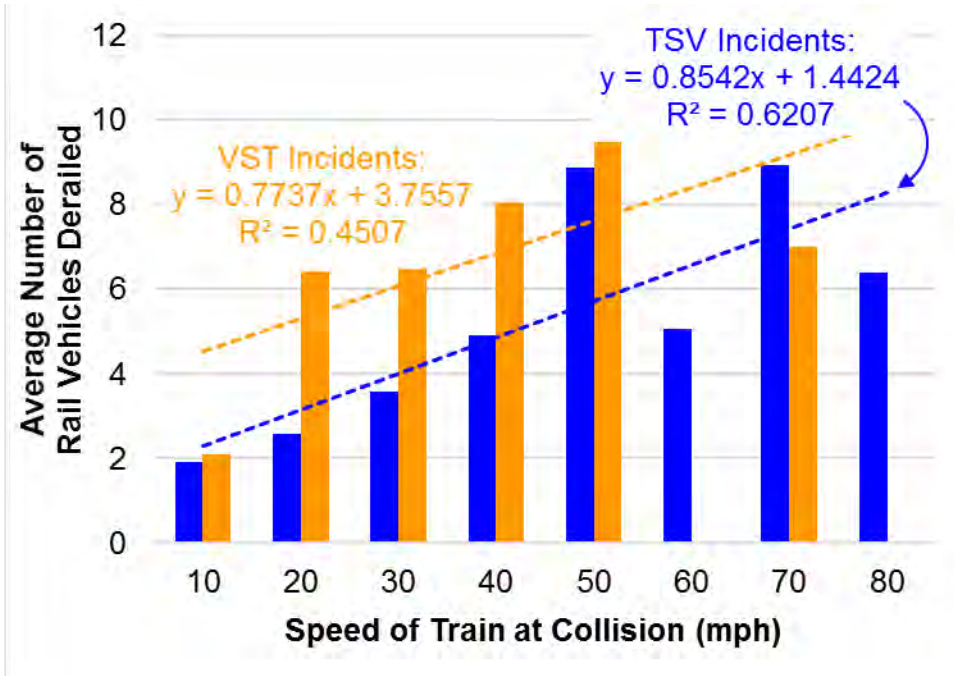


Figure 4.12: Relationship between average number of rail vehicles derailed per incident and speed of train at collision. Orange bars represent VST incidents and blue bars represent TSV incidents. Note: Analysis is for large highway vehicles only, as no relationship was observed for small highway vehicles.

4.3.7. Train Weight

Train weight is another physical factor that could affect incidents. Information on the weight of rolling stock involved in the incident is not provided in the HRA or REA databases, though some information about the "first [rail vehicle] involved" is reported, namely its reporting mark. The reporting mark is a unique identifier that can be cross-referenced with other sources to provide additional information about the rail vehicle.

4.3.7.1. Weight of First Railcar Involved

Due to the complexity of gathering the weight data, the rolling stock weight variable was approached simultaneously in two ways. Data on weight was manually gathered for approximately 1,000 HRA/REA records based on reporting mark. Two sources – the Universal

Machine Language Equipment Register (UMLER), and the Official Railway Equipment Register (ORER) – provide information about the loaded and tare weights of rail vehicles; however, for most records in the HRA/REA databases these sources must be cross-referenced manually, especially for rail vehicles that may have been scrapped after their involvement in an incident. The UMLER database is online and easily-referenced, but only has data for rail vehicles currently in service. Therefore, data for older rail vehicles must be sourced by hand from print books such as the ORER, or websites in the case of privately-held railroad equipment. Information about rail equipment can often be found on railroad photography websites (such as railroadpictures.net) when other sources fail.

Due to a combination of reporting errors and gaps in the weight databases, roughly 5% of records could not be matched to weight data and were discarded. The discarded records were predominantly related to the oldest incidents (those between 1991 and 1995) creating a time bias in the completeness of the weight data. The disadvantage of this method of accounting for rail vehicle weight is the time required. It was infeasible to collect data for all the records in the database, so instead a random selection technique was used to create a sample dataset of 1,000 records.

The effect of rail vehicle weight differed for passenger and freight trains (Table 4.1). For freight trains, average weight varied significantly between derailment and non-derailment incidents (2-tailed t-test, p-value < 0.0001), with derailments more likely if lighter rail vehicles were involved. In contrast, passenger trains showed no significant difference in average weight between derailment and non-derailment incidents (2-tailed t-test, p-value $= 0.7717$). Note that there are fewer records for passenger trains than freight trains, which could contribute to the lack of significance. Additionally, passenger equipment weights vary widely in the U.S.

Table 4.1: Effect of Rail Vehicle Weight on Derailment Likelihood

	Passenger		Freight	
	Derailment	Non-Derailment	Derailment	Non-Derailment
Average Weight (lbs)	245,121	248,058	330,579	359,595
Standard Deviation	45,177	61,214	84,474	63,005
N	47	92	194	642
t-test p-value	0.7717		<0.0001	

4.3.7.2. Equipment Type

While the exact-weight analysis was being developed, another, simpler method for quantifying rail vehicle weight was developed. Though the weight of rail vehicles varies widely, in general, freight rail vehicles are heavier than passenger vehicles. Additionally, locomotives are typically heavier than railcars. Therefore, if the type of rail vehicle (freight or passenger, locomotive or railcar) first involved in the incident could be determined, an ordinal variable with four values could be created representing rail vehicle weight to approximate the weight effect. Further differentiation between loaded and unloaded freight cars would have been ideal, since the former are heavier and the latter are lighter than passenger cars. While the loading condition of the railcar is reported in the REA database, this information is not included in the HRA database. Since most grade crossing incidents do not exceed the REA reporting threshold, this information does not exist for most incidents and therefore could not be analyzed.

To determine the type of rail vehicle involved in the incident, some assumptions were made based on the data in the HRA database. First, the field TYPEQ, which identifies the type of equipment in the train consist at the time of the incident, was used to determine whether a rail vehicle was passenger or freight. There are 10 types of consist specified, and these were divided into two groups based on whether they were more like passenger cars or freight cars. For example, category 9 is defined as a maintenance or inspection car. Such rail vehicles are often adapted passenger cars or various high-rail vehicles which are motor vehicles adapted to operate on railroad tracks. These were collectively classified as "passenger". A complete list of the TYPEQ categories and freight/passenger sorting is shown in Appendix B.

It was then necessary to determine if a rail vehicle was a locomotive or railcar. Some categories only apply to one or the other. For example, incidents in category 6 – cut of cars – certainly involve a railcar. However, for some categories, such as "freight train", the individual rail vehicle involved in the incident could be either a locomotive or a railcar. Therefore, it was necessary to use another field in the database to determine if a rail vehicle was a locomotive or a railcar. I compared two fields – RRCAR (indicating the position-in-train of the first railcar involved in the incident) and NBRLOCOS (indicating how many locomotives were in the train consist) – and assumed that the struck rail vehicle was a locomotive if the value of RRCAR was less than the value of NBRLOCOS. This is an oversimplification, especially in the case of freight trains, which may use distributed power, meaning locomotives are placed elsewhere in the consist besides the front. However, it is not possible to determine from the HRA database how many of the consist's locomotives were at the front of the train.

Comparing the category distributions for derailments and non-derailments shows that derailments are not occurring at the same rate for all equipment types (χ^2 test with 3df $p<0.0001$). For incidents in which the struck rail vehicle was a freight locomotive, derailments occurred only 0.73% of the time, whereas 16.67% of incidents involving passenger railcars resulted in derailments (Table 4.2).

Table 4.2: Number and Percentage Derailment of Incidents by Equipment Type

	Derailment[7]	Non-Derailment	Percent Derailments
Freight Locomotive	251	34,529	0.73%
Passenger Locomotive	50	3,671	1.36%
Freight Railcar	80	4,075	1.96%
Passenger Railcar	15	90	16.67%

4.3.8. Crossing Angle

The last physical factor, crossing angle, measures the angle at which the train and highway vehicle collide. While there is a variable describing crossing angle (XANGLE) in the FRA's Grade Crossing Inventory (GCI), this variable is categorical, not continuous. Grade crossing angle is classified into three categories based on the smallest (acute) angle of incidence (Table 4.3).

Table 4.3: Grade Crossing Inventory XANGLE Value and Actual Angle Ranges

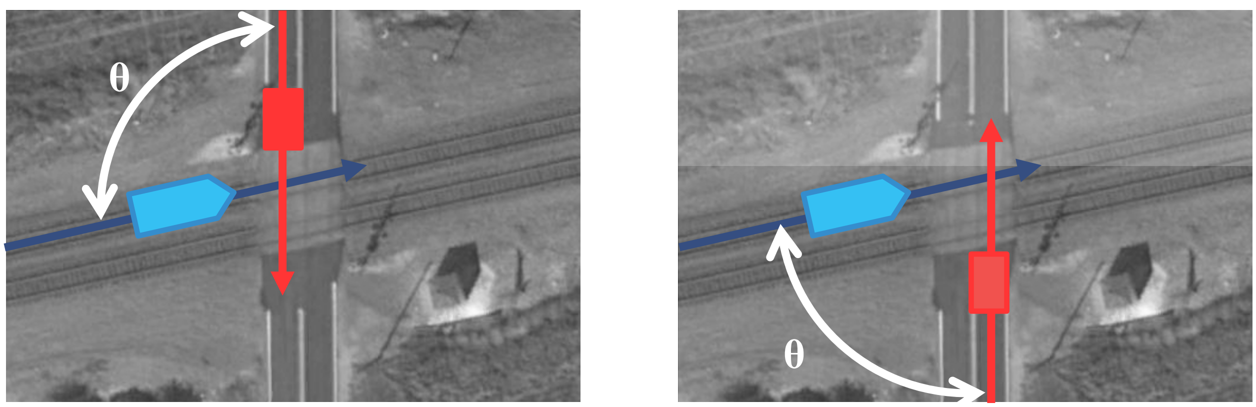
XANGLE Value	Acute Angle Range	Obtuse Angle Range
1	0-29°	180-151°
2	30-59°	150-121°
3	60-90°	120-90°

Consequently, exact collision angle is not available from the GCI. For example, a crossing might have XANGLE = 1, indicating that the smallest crossing angle is between 0° and 29°. This means the actual angle of collision could be between 0° and 29° or between 151° and 180° (Figure 4.13). The collision angle could affect the closing velocity between the train and

[7] Three derailments involved maintenance-of-way equipment (TYPEQ=A) and were excluded.

highway vehicle since speed is a vector. Crossing angle could also have an influence on derailment likelihood since it may affect the interaction of the train's wheels with the rail. Therefore, it was hypothesized that these two cases could have different implications for the physics of the collision.

For a data set of 3,062 unique mainline freight train grade crossing collisions, the latitude and longitude of the crossing, as well as the travel direction of the train and the car, were determined from HRA and GCI records. Google Earth was used to analyze aerial photos of each grade crossing, and the exact angle of collision, θ, between the train and the highway vehicle was measured (Figure 4.13).



(a) (b)

Figure 4.13. Diagram showing angle of collision, θ. Notice that although the crossing is the same in both collisions, the angle of collision is greater in (a) than in (b) because of the direction the highway vehicle is traveling.

Approximately one third of the data points were excluded from the analysis because of errors or omissions in the data. For example, in many instances the longitude and latitude of the crossing were missing or inexact, making it impossible to locate the crossing. Additionally, there were reporting errors in travel direction that made the data ambiguous. For example, the highway vehicle's direction of travel would be reported as "west" when the road it was traveling on was oriented north-south. (It was sometimes possible to determine by zooming out that the road was generally oriented east-west, in which case it was assumed that this was the correct direction.) Complete data were collected for 2,064 entries.

The crossing angle data were divided according to whether a derailment occurred and the two distributions were compared. Angle of collision is normally distributed for both samples. The mean angle of the two samples was approximately the same. Derailments occurred over a slightly smaller range of values (Table 4.4).

Table 4.4: Summary Statistics Divided by Derailment Occurrence

Incident Type	N	Mean	Standard Deviation	Standard Error	Min Value	Max Value
Non-Derailment	1,424	87.68°	20.95	0.5551	15°	155°
Derailment	220	90.95°	23.19	1.5633	25°	160°

A t-test was performed to determine if the distribution of crossing angles was different for incidents in which derailments did or did not occur. The results indicated that the means of the two samples were unequal ($Pr > F = 0.0339$; $\alpha=0.05$) and the difference between the two populations is statistically significant. Specifically, derailments were slightly more likely to occur at higher collision angles.

Ultimately, it is difficult to determine exactly what effect crossing angle has on derailment likelihood. If the train and highway vehicle were both rigid bodies on smooth surfaces, then the effect would follow logically from basic energy transfer equations. However, they are not rigid bodies, so they will absorb some portion of the energy involved in the collision. Combined with the interaction that exists between rails and train wheels due to the wheels' flanges, the effect of collision angle is probably mitigated. The best way to study this component would be through detailed computer modeling of the collision.

4.3.9. Other Variables

Various railroad practitioners who have experience with grade-crossing-collision-caused derailments have suggested other variables that might affect derailment occurrence. Two variables that are consistently mentioned are the curvature and gradient of the track at the crossing. It has been suggested that because curved track is generally superelevated, a collision occurring on a curve might be more likely to result in derailment if the train is struck from the high side of the curve (outside rail) compared to the low side (inside rail). It has also been suggested that grade could play a role in derailment probability. Trains traveling downgrade at the time of a collision might have more momentum because of the effect of gravity, and therefore have greater derailment potential. Neither grade nor curvature are provided in the FRA Grade Crossing Inventory or any other feasibly referenced source, therefore I have not investigated the effect of these factors. These data could be obtained from railroad track charts by future researchers.

4.4. DISCUSSION

This chapter examined each of the physical factors identified in Chapter 1. Large highway vehicles such as trucks appear to be about three times more likely to cause grade crossing derailments than small vehicles. Large vehicles are, by a considerable margin, involved in a disproportionately greater number of derailments considering the number of incidents they are involved in (Figure 4.7). This result is not surprising; in cases where a train hits a large, heavy highway vehicle its greater mass may make it more likely to dislodge the train from the tracks, and it is also capable of absorbing more of the train's momentum causing severe run-in and possible jack-knifing of the train. A smaller, lighter vehicle is more likely to be pushed down the tracks allowing the train to lose speed more gradually.

Highway vehicle size did not appear to have much effect on derailment severity. While the most severe incidents – those resulting in the derailment of more than 25 cars and locomotives – were generally caused by trucks, they accounted for only 3% of all incidents. Statistical analysis found no significant difference between the two severities.

Impact velocity of both the highway vehicle and the train were found to influence derailment rate. Derailments tended to occur at higher vehicle speeds regardless of incident type. Also, derailments tended to occur at higher train speeds for both TSV and VST incidents though the trend is stronger for TSV incidents. That higher speeds result in more derailments is not surprising given that the energy involved in high speed collisions is greater than other collisions, making the train more likely to leave the tracks.

Additionally, some interesting patterns were evident in the velocity study. For the "train striking vehicle" category, about 37% of these incidents occurred between a train and a vehicle that was stationary on the tracks (with a speed of 0 mph). 31% of all derailments for this category occurred with a stationary vehicle. Examination of the narrative fields for these incidents showed that at least 40% were caused by a truck being stuck on a crossing and unable to move in time. This suggests that further efforts to modify crossing geometry so that trucks are less likely to get stuck would yield benefits. In addition, more frequent crossing inspections combined with careful route planning could prevent trucks from becoming stuck by directing them to more appropriate routes that properly account for their under-truck clearance.

The rail vehicle weight study indicated that derailment likelihood varies according to equipment type, with freight locomotives being the least likely to derail, and passenger railcars being most likely. An analysis using actual rail vehicle weights was also conducted, but the results were less conclusive, probably due to the size variance of the dataset. Exact rail vehicle weight data are difficult to obtain given current incident reporting requirements.

The collision angle analysis showed that derailments are slightly more likely to occur at higher collision angles. All else constant, the effective collision force will increase with increased collision angle. One caveat is that this analysis was conducted using a sample size of only 2,000 incidents instead of the full dataset. This was unavoidable due to the complexity of gathering exact collision angle information, as well as inaccuracy in the GCI database. While there were enough randomly-selected records to provide statistically significant results, it is possible that some bias was introduced into the analysis through the selection process. For example, some states might provide more accurate latitude/longitude coordinates than others, meaning they would have more crossings that could be identified in Google Earth during the data collection process. Combined with the fact that some states are making efforts to eliminate non-right-angle crossings, this could create a sample that is not completely representative. However, regional variations should not affect incident physics, so this is less likely to be a problem.

4.5. CONCLUSIONS

This analysis considered mainline, REA-reportable incidents from the years 1991 to 2010, using a large dataset that was developed in order to obtain a robust sample size and greater statistical power. The major physical factors affecting grade crossing incidents and their effect on derailment rate and severity were investigated. The purpose was to identify which factors have an effect on these metrics and to facilitate development of a multivariate model to predict derailment risk at a given grade crossing. The results show that vehicle size has a strong effect on derailment rate, but little effect on derailment severity. Vehicle and train speed at collision also influence derailment rate. Additionally, derailment severity increases with increased train speed, provided the incident involves a large highway vehicle.

CHAPTER 5: STATISTICAL MODELING OF FREIGHT TRAIN DERAILMENTS

5.1. INTRODUCTION

This chapter describes development of a statistical model to estimate the conditional probability of a grade crossing collision resulting in a freight train derailment, based on a variety of characteristics of the collision. Modeling began with incidents involving only freight trains because freight and passenger trains have different characteristics and consequences. Additionally, while grade crossing incidents involving passenger trains are approximately two times more likely than freight trains to result in derailment, freight train data are much more numerous. As will be discussed in Chapter 6, development of a separate passenger model was ultimately not possible. Development of the freight train models led to interesting observations about how best to create suitable statistical models, which in turn helped with creation of the final model presented in Chapter 6.

5.2. DATASET

The analysis in this chapter combines data from the Rail Equipment/Accident (REA), Highway-Rail Grade Crossing Accident (HRA) and Grade Crossing Inventory (GCI) databases. Data for all U.S. freight railroads during the 20-year period 1991 through 2010 were used. The HRA database contains factors useful for developing a predictive model; however, it does not provide a means of identifying incidents resulting in derailment, which are found in the REA database. Additional factors, including information about the characteristics of the crossing, come from the GCI. This study focused on grade crossing incidents occurring on mainline track and sidings because these accounted for approximately 88% of all incidents. The databases were combined as explained in Chapter 3.

Since different data are included in the HRA and REA databases, three different datasets were used to help develop the statistical model (Table 5.1). Dataset “A” includes all unique incidents reported in the HRA database that involved freight equipment on mainline track. Dataset “B” includes those incidents that were reported in both the HRA and REA databases. These incidents are assumed to be the most severe in regards to damaging rail equipment, track and structures, since they exceeded the reporting threshold for the REA database. Dataset “C” is a subset of dataset “B” that includes all grade crossing derailment events and a randomly selected subset of non-derailment events.

Table 5.1: Summary of Datasets Used in Model Development

Dataset	Derailments	Non-Derailments	Total Events
A	312	43,326	43,638
B	312	1,934	2,246
C	312	624	936

5.3. METHODOLOGY

The goal of this chapter is to develop a statistical model to predict the conditional probability of a highway-rail grade crossing collision resulting in derailment of a freight train given that a collision has occurred. The model has a binary response variable: either there is a derailment or there is not. The input variables to the model were the physical factors identified in Chapters 1 and 4: highway vehicle size, highway vehicle speed, train speed, crossing angle, and incident type. Train mass was not considered in this chapter. The input variables were binary, categorical, or continuous in nature (Table 5.2).

Table 5.2: Definition of Model Variables

Variable Name	Definition (FRA 2011)	Variable Type	Range of Values
VEHSPD	Highway vehicle estimated speed in mph	Continuous	Range*: 0-105 mph Average*: 10.50 mph Standard Deviation*: 13.57
TRNSPD	Train speed in mph	Continuous	Range*: 0-80 mph Average*: 31.45 mph Standard Deviation*: 15.58
LGVEH	Was a large highway vehicle involved?	Binary	N if no; Y if yes
TRNSTK	Did train strike highway user?	Binary	N if highway user struck train; Y if train struck highway user
TRKCLAS	FRA track class	Categorical	0-9 (0 represents X)
WARNSIG	Crossing warning interconnected with highway signals	Categorical	1 if yes; 2 if no; 3 if unknown
VIEW	Was the driver's view of the track obstructed?	Binary	N if not obstructed; Y if obstructed
PUBLIC	Did the collision occur at a public crossing?	Binary	Y if public; N if private
XTYPE	Type of warning device at crossing	Categorical	1: gates 2: active (excl. gates) 3: passive 4: other 5: none

* Note: Statistics are for dataset A. Sets B and C have similar statistics.

Track class, crossing visibility, type of warning device and accessibility of crossing (public or private) were also considered, since they are characteristics that appear in many incident prediction models. Second-order interaction and polynomial effects were considered for the continuous variables.

While the quality of the databases is generally quite good, there were some data points that were internally inconsistent or had empty fields. Additionally, some fields were re-coded based on groupings used in the model. A detailed procedure for data clean-up can be found in Appendix A.

The model was developed using the LOGISTIC procedure in the Statistical Analysis Software (SAS) program. This procedure uses the method of maximum likelihood to fit a linear logistic regression model to binary response data (SAS Institute, 2013). In this way, the relationship between certain explanatory variables and response variables can be analyzed. For each collision, the output of the model is a value between 0 and 1 representing the probability of a derailment occurring.

The SAS LOGISTIC procedure has four effect selection techniques: forward selection, backwards elimination, stepwise selection and best subset selection (SAS Institute, 2013). All four techniques were evaluated and it was found that, for this dataset, stepwise selection was best for model formulation. In stepwise selection, the variable having the strongest influence on the response variable is added at each step. Variables are not added if they are found to have an insignificant influence on the model. At each step, the procedure tests the influence of including each factor by performing a "goodness-of-fit" test. It also examines the factors that have already been added and removes any that are found to no longer have a significant effect. Additionally, SAS indicates any co-linear variables to the user, as highly correlated variables can impair the ability of SAS to identify the best variable for inclusion in the model.

The model produced by SAS LOGISTIC identifies the probability of a derailment having occurred as the result of a given grade crossing collision over the past 20 years. Dick et al. (2001) define this as a "retrospective" model, as opposed to a "prospective" model. The retrospective model makes predictions about past events using a subset of the data, consisting of 33% derailments and 67% non-derailments. The output of this retrospective model must be calibrated to more accurately represent the probability of a derailment occurring in the overall population. This adjusted "prospective" model could be used to identify grade crossings with a greater likelihood of derailment.

Several models were produced using the LOGISTIC procedure. Initially, models were produced using dataset B; however, these models proved to be inaccurate, regardless of whether interaction effects were included in the LOGISTIC procedure. It was hypothesized that, since there are six times more non-derailment events than derailment events, the models were forced to predict non-derailment events more correctly at the expense of predicting derailment events. In this way, the model predicted a large percent of all events correctly, but had poor fit because it failed to predict most derailment events, which are of the greatest interest in this study.

This problem can be remedied by using a modified form of logistic regression referred to as "rare events logistic regression" (RELR) (King and Zeng, 2001; van den Eeckhaut et al., 2006). RELR corrects for the disproportionate number of non-events by selecting a random subset of non-events, which are equal to 1 to 5 times the number of events. Therefore, dataset C was created, containing a number of randomly-selected, non-derailment events equal to twice the number of derailment events.

5.4. RESULTS

5.4.1. Unified Model

5.4.1.1. Retrospective Derailment Model

Four subsets of dataset B were created (C1 through C4) using all 312 derailment events and 624 randomly-selected non-derailments. The non-derailments were different in each of the four datasets. Based on dataset C3, the following model was developed:

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

$$x = -0.6001 + \begin{cases} 0, & LGVEH = Y \\ -0.4106, & LGVEH = N \end{cases} + \begin{cases} 0, & TRNSTK = Y \\ 0.3822, & TRNSTK = N \end{cases} + 0.0316\, VEHSPD - 0.0141\, TRNSPD$$

(0.1059) (0.1019) (0.00530) (0.00533) (s.e)

where VEHSPD, TRNSPD, TRNSTK and LGVEH are described in Table 5.2.

This model provided the best fit to the data, with a Hosmer-Lemeshow (HL) goodness-of-fit test result of 0.5771, indicating that the model performs slightly better than random guessing. This model also has a suitable ability to discriminate between derailment and non-derailment events, as measured by the Receiver Operating Characteristics (ROC) curve, with an area under the ROC curve of 0.7006 (Figure 5.1).

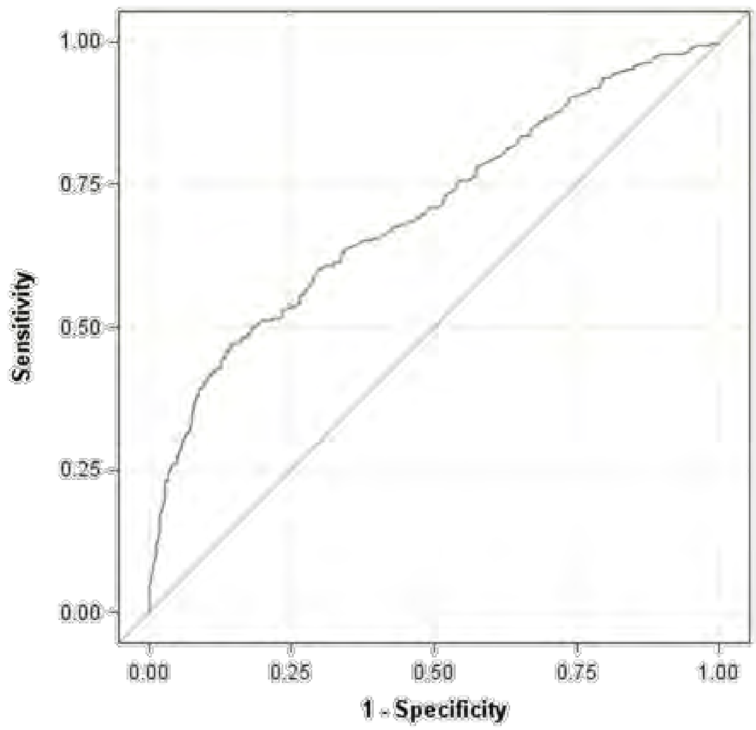


Figure 5.1 ROC Curve for retrospective model. Area under the curve is 0.7006.

Additional performance statistics for this model are presented in Table 5.3. For these values, the threshold value for predicting a derailment was a p value of 0.3. If the calculated value of *p* for a data point was greater than 0.3, it was classified as predicting a derailment, and if it was less than 0.3, it was classified as predicting no derailment.

Table 5.3: Performance Statistics for Retrospective Model

Statistic	Cases
Percent Correct	66.2%
Sensitivity	59.6%
Specificity	69.4%

In this model, the intercept term (*b = -0.6001*) is based on the average probability of a derailment for dataset C. While the factor coefficients from the small dataset are equally valid for the large dataset, the intercept term needs to be adjusted in the prospective model in order to account for the average rate of derailment in the actual population (Scott and Wild, 1986). The next two sections consider a prospective model based on dataset A.

5.4.1.2. Prospective Derailment Model for All Grade Crossing Collisions (Dataset A)

In this section, the retrospective model is adjusted to reflect the derailment rate for all grade crossing collisions. As explained previously, in order to be able to predict the likelihood of a highway-rail grade crossing collision resulting in a derailment, the retrospective model developed above must be altered to become a predictive model. This is accomplished by altering the intercept term to account for the 20-year average likelihood of a derailment occurring. For dataset A, the average derailment likelihood, $p_{avg(A)}$ can be calculated as

$$p_{avg(A)} = \frac{312\ derailments}{43{,}638\ total\ events} = 0.0071$$

The intercept term is then modified to account for $p_{avg(A)}$ using the log-odds operator (Dick 2001).

$$b_A = b + \ln\left(\frac{p_{avg(A)}}{1 - p_{avg(A)}}\right)$$

$$b_A = -0.6001 + \ln\left(\frac{0.0071}{1 - 0.0071}\right) = -5.5349$$

Using the modified intercept term adjusts the probabilities predicted by the model to more accurately reflect the actual observed rate of derailments. Based on dataset A, the following model was produced:

$$p_A = \frac{1}{e^{-x_A} + 1}$$

$$x = -5.5349 + \begin{cases} 0, & LGVEH = Y \\ -0.4106, & LGVEH = N \end{cases} + \begin{cases} 0, & TRNSTK = Y \\ 0.3822, & TRNSTK = N \end{cases} + 0.0316\, VEHSPD - 0.0141\, TRNSPD$$

where VEHSPD, TRNSPD, TRNSTK and LGVEH are described in Table 5.2.

An ROC curve was generated by analyzing dataset A with equation p_A (Figure 5.2). The area under the ROC curve was 0.7551, which is considered acceptable discrimination.

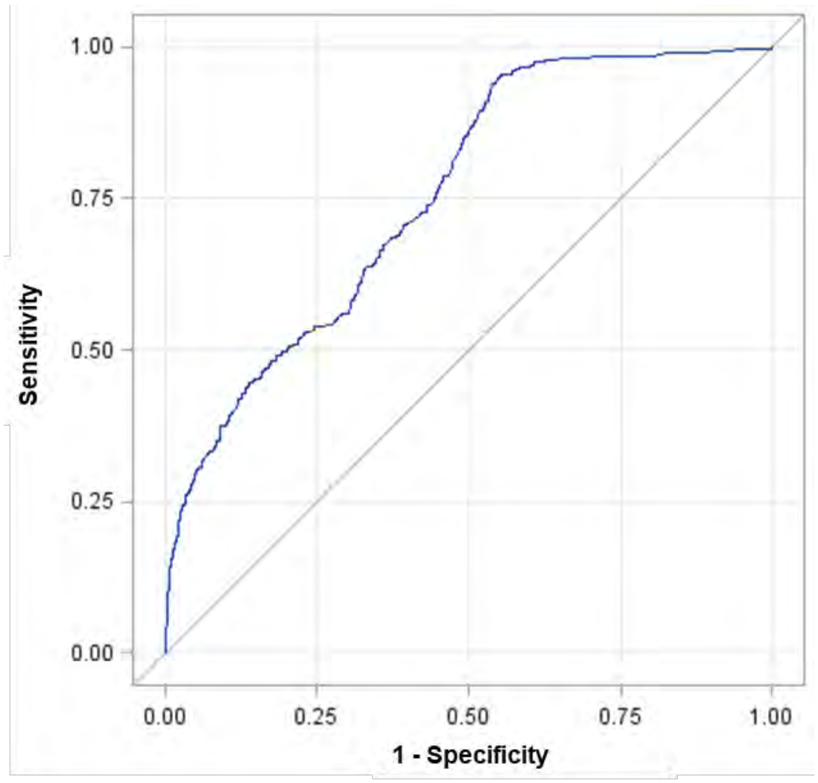


Figure 5.2: ROC Curve for dataset "A". Area under the ROC curve is 0.7551.

5.4.2. Split Model

While the model developed in Section 5.4.1 is reasonably effective at describing the factors that determine whether or not a grade crossing incident will result in derailment, there is still some unaccounted-for variance. As explained in Chapter 1, the incident type (train striking vehicle or vehicle striking train) was expected to play a key role in the statistical model. Incident type was selected as a factor, but there were no interaction effects with the other factors.

Hypothetically, since the physical effects involved in each incident type are probably different, the single model might be having a difficult time separating them, even using interaction effects. Therefore, the dataset was split into two parts based on incident type and developed as two separate models. The methodology for this is the same as described in Section

4.1, starting with a retrospective model using a small dataset, and expanding to a prospective model using the full dataset (Table 5.4).

Table 5.4: Summary of Datasets Used in Model Development

Dataset	Number of Derailments	Number of Non-Derailments	Total Number of Events
AVST	110	8,639	8,749
ATSV	202	34,687	34,889
CVST	110	220	330
CTSV	202	404	606

Dataset AVST represents all freight VST incidents reported in the HRA database. Dataset ATSV represents all TSV freight incidents reported in the HRA database. Similarly, dataset CVST represents all VST derailments and a number of randomly selected VST non-derailments equal to twice the number of derailments. Dataset CTSV represents all TSV derailments and a number of randomly selected TSV non-derailments equal to twice the number of derailments. CVST and CTSV were used to develop logistic regression models. As with the unified model, four of each "C" dataset were developed and tested to ensure that any observed trends were not artifacts of the specific selected records. From these, the statistical model with the best fit characteristics was selected.

5.4.2.1. Retrospective Derailment Models

Based on dataset CVST, the following model was developed for incidents where the vehicle struck the train.

$$p_{CVST} = \frac{1}{e^{-x_{CVST}} + 1}$$

$$x_{CVST} = -2.0403 + \begin{cases} 0, & LGVEH = Y \\ -1.5044, & LGVEH = N \end{cases} + 0.00101\, VEHSPD^2$$

(0.2558) (0.1853) (0.000156) (s.e)

where VEHSPD and LGVEH are described in Table 5.2.

This model provided the best fit to the data, with an HL test result of 0.9920. The area under the ROC curve for this result is 0.8871 (Figure 5.3). Additional performance statistics for this model are given for a derailment threshold p value of 0.3 (Table 5.5).

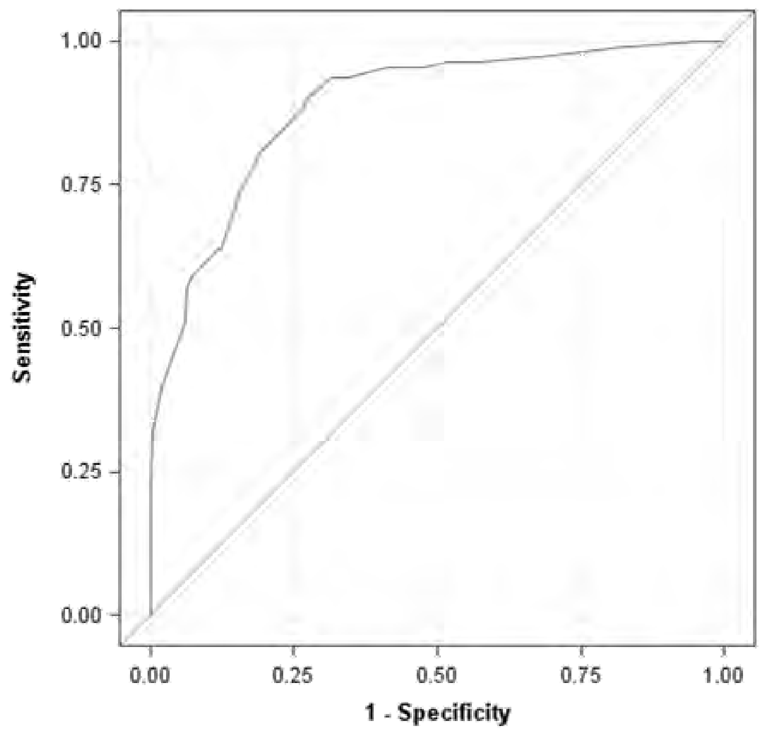


Figure 5.3: ROC Curve for retrospective model. Area under the curve is 0.8871.

Table 5.5: Performance Statistics for Retrospective Model "CVST"

Statistic	Cases
Percent Correct	78.2%
Sensitivity	90.0%
Specificity	72.3%

Based on dataset CTSV, the following model was developed for incidents where the vehicle struck the train.

$$p_{CTSV} = \frac{1}{e^{-x_{CTSV}} + 1}$$

$$x_{CTSV} = -2.0330 + \begin{cases} 0, & LGVEH = Y \\ -1.8687, & LGVEH = N \end{cases} + 0.0166\,TRNSPD$$

(0.3239) (0.1712) (0.00766) (s.e)

where TRNSPD and LGVEH are described in Table 5.2.

This model has an HL goodness-of-fit test result of 0.8110. The area under the ROC curve for this result is 0.8422 (Figure 5.4). Additional performance statistics for this model are given for a derailment threshold *p* value of 0.3 (Table 5.6)

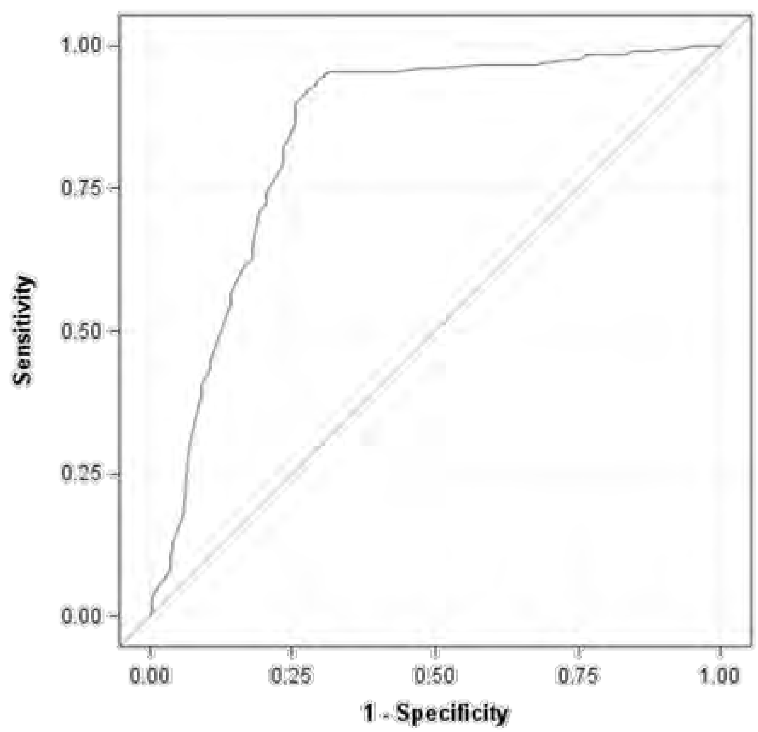


Figure 5.4: ROC Curve for retrospective model. Area under the curve is 0.8422.

Table 5.6: Performance Statistics for Retrospective Model "CTSV"

Statistic	Cases
Percent Correct	77.9%
Sensitivity	95.0%
Specificity	69.3%

In these models, the intercept terms ($b_{CVST} = -2.0403$; $b_{CTSV} = -2.0330$) are based on the average probability of a derailment for datasets CVST and CTSV. The next section develops the prospective models based on datasets AVST and ATSV.

5.4.2.2. Prospective Derailment Model (Datasets AVST and ATSV)

In this section, the retrospective models are adjusted to reflect the derailment rate for all grade crossing collisions of a given incident type. This is accomplished by altering the intercept term to account for the 20-year average likelihood of a derailment occurring. For dataset AVST, the average derailment likelihood, $p_{avg(AVST)}$ can be calculated as

$$p_{avg(AVST)} = \frac{110\ derailments}{8{,}749\ total\ events} = 0.0126$$

The intercept term is then modified to account for $p_{avg(AVST)}$ using the log-odds operator.

$$b_{AVST} = b_{CVST} + \ln\left(\frac{p_{avg(AVST)}}{1 - p_{avg(AVST)}}\right)$$

$$b_{AVST} = -2.0403 + \ln\left(\frac{0.0126}{1 - 0.0126}\right) = -6.4039$$

Using the modified intercept term adjusts the probabilities predicted by the model to more accurately reflect the actual observed rate of derailments. Based on dataset AVST, the following model was produced:

$$p_{AVST} = \frac{1}{e^{-x_{AVST}} + 1}$$

$$x_{AVST} = -6.4039 + \begin{cases} 0, & LGVEH = Y \\ -1.5044, & LGVEH = N \end{cases} + 0.00101\, VEHSPD^2$$

where VEHSPD and LGVEH are described in Table 5.2.

An ROC curve was generated by analyzing dataset AVST with equation p_{AVST}. The area under the ROC curve was found to be equal to 0.8800, which is considered to be good discrimination.

For dataset ATSV, the average derailment likelihood, $p_{avg(ATSV)}$ can be calculated as

$$p_{avg(ATSV)} = \frac{202\ derailments}{34{,}889\ total\ events} = 0.0058$$

The intercept term is then modified to account for $p_{avg(ATSV)}$ using the log-odds operator.

$$b_{ATSV} = b_{CTSV} + \ln\left(\frac{p_{avg(ATSV)}}{1 - p_{avg(ATSV)}}\right)$$

$$b_{ATSV} = -2.0330 + \ln\left(\frac{0.0058}{1 - 0.0058}\right) = -7.1789$$

Using the modified intercept term adjusts the probabilities predicted by the model to more accurately reflect the actual observed rate of derailments. Based on dataset ATSV, the following model was produced:

$$p_{ATSV} = \frac{1}{e^{-x_{ATSV}} + 1}$$

$$x_{ATSV} = -7.1789 + \begin{cases} 0, & LGVEH = Y \\ -1.8687, & LGVEH = N \end{cases} + 0.0166\, TRNSPD$$

where TRNSPD and LGVEH are described in Table 5.2.

An ROC curve was generated by analyzing dataset ATSV with equation p_{ATSV}. The area under the ROC curve was found to be equal to 0.8360, which is considered to be good discrimination.

5.5. DISCUSSION

5.5.1. Interpretation of Model Terms – Unified Model

The model presented in Section 5.4.1 contains four terms that indicate the effects of different vehicle and incident characteristics. Out of the nine variables provided in the dataset, the SAS LOGISTIC procedure selected four variables as providing the best fit model. Table 5.7 summarizes the order that the terms were added to the model. It is interesting to note that no term was added and then removed in a subsequent step.

Table 5.7: Summary of of Model Variable Selection

Step	Term Added	Term Removed	Chi-Square
1	VEHSPD	--	109.90
2	LGVEH	--	16.14
3	TRNSTK	--	16.75
4	TRNSPD	--	7.11

The first term in the model, *0.0316 VEHSPD*, indicates that the speed of the vehicle at collision affects derailment likelihood. The probability of derailment increases with higher vehicle speed.

The second term in the model, $\begin{cases} 0, & LGVEH = Y \\ -0.4106, & LGVEH = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle such as a car, motorcycle or pickup-truck (*LGVEH* = *N*) then this term assumes a value of -0.4106; if the highway user is a large vehicle such as a tractor-semi-trailer or a straight truck (*LGVEH* = *Y*) then the term disappears and probability increases. This means that, all else equal, a collision where the highway user is a large vehicle is more likely to result in a derailment.

The third term in the model, $\begin{cases} 0, & TRNSTK = TSV \\ 0.3822, & TRNSTK = VST \end{cases}$, shows that the circumstances of the collision have an important effect. If the highway user strikes the train (*TRNSTK = VST*) then this term assumes a value of 0.3822; if the train strikes the highway user (*TRNSTK = TSV*) then the term disappears and probability decreases. This means that, all else equal, a collision where the highway vehicle strikes the train is more likely to result in a derailment.

The final term in the model, *-0.0141 TRNSPD*, shows that the speed of the train at collision has an effect on derailment likelihood. As train speed increases, the probability of derailment decreases. This is the opposite of the trend observed in the univariate analysis in Chapter 4 and will be discussed further in Section 5.5.2 below.

Since both TRNSTK and LGVEH are binary variables, it is possible to directly compare their coefficients. These coefficients suggest that the model is slightly less sensitive to incident type than to highway vehicle type. Similarly, the coefficients of VEHSPD and TRNSPD can be compared because both are continuous variables with similar ranges. The model is slightly more sensitive to vehicle speed than to train speed.

5.5.2. Interpretation of Model Terms – Split Model

The models presented in Section 5.4.2 each contain two terms that indicate the effects of different vehicle and incident characteristics. Out of the nine variables provided in the dataset, the SAS LOGISTIC procedure selected three variables between the two models. Tables 5.8 and 5.9 summarize the order that terms were added to the models. No term was added to the model and then removed in a subsequent step.

Table 5.8: Summary of of Model Variable Selection – CVST

Step	Term Added	Term Removed	Chi-Square
1	LGVEH	--	101.70
2	VEHSPD^2	--	54.56

The first term in the model for VST incidents, $\begin{cases} 0, & LGVEH = Y \\ -1.5044, & LGVEH = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle (*LGVEH = N*) then this term assumes a value of -1.5044; if the highway user is a large vehicle (*LGVEH = Y*) then the term disappears and probability increases. This means that, all else equal, a collision where the highway user is a large vehicle is more likely to result in a derailment.

The second term in the model, $0.00101\ VEHSPD^2$, indicates that the speed of the vehicle at collision affects derailment likelihood. As vehicle speed increases, the probability of derailment also increases.

Table 5.9: Summary of of Model Variable Selection – CTSV

Step	Term Added	Term Removed	Chi-Square
1	LGVEH	--	223.51
2	TRNSPD	--	4.75

The first term in the model for TSV incidents, $\begin{cases} 0, & LGVEH = Y \\ -1.8687, & LGVEH = N \end{cases}$, indicates that the type of highway vehicle involved in the incident affects derailment likelihood. If the highway user is a small vehicle (*LGVEH = N*) then this term assumes a value of -1.8687; if the highway user is a large vehicle (*LGVEH = Y*) then the term disappears and probability increases. This means that, all else equal, incidents involving large vehicles are more likely to result in a derailment.

The second term in the model, *0.0166 TRNSPD*, shows that the speed of the train at collision has an effect on derailment likelihood. As train speed increases, the probability of derailment increases, consistent with the work presented in Chapter 4. This is the opposite of the trend observed in the unified model, likely because the unified model attempted to predict two different trends with only one variable.

As confirmation of the value of the split models, it is interesting to note that each has two factors that are logical considering the differences in physical characteristics of VST and TSV incidents. For VST incidents, the key factors are vehicle size and speed. For TSV incidents, the key factors are vehicle size and train speed.

5.5.3. Discussion of Threshold Value

In Section 5.4, receiving operator characteristic curves were shown for each of the models as a way of demonstrating the model's ability to discriminate between derailment and non-derailment events. A more traditional way of demonstrating discrimination is to show the sensitivity (proportion of correctly identified derailments), specificity (proportion of correctly identified non-derailments) and correct predictions for a model, at a given threshold value, as shown in Tables 5.3, 5.5 and 5.6. However, the predictive accuracy of the model varies widely based on the threshold value selected. This threshold value is often based on the observed likelihood of the event, but can also be decided on as a matter of policy. For example, Table 5.3 gives performance statistics for the retrospective model at a threshold value of 0.30. For this cut point the model has a percent correct, sensitivity and specificity all between 60-70%. It is possible to increase the percentage of correct predictions and the specificity by raising the threshold; however, this will reduce the sensitivity. In the case of derailment prediction, as a matter of policy it might be preferable to overestimate risk instead of underestimating it. In this case, it may be desirable to select a threshold with greater sensitivity, because the model will then correctly identify a larger percentage of derailments.

The models presented here are capable of differentiating between collisions that are likely to result in derailments, and those that are not. This is an additional factor that railroads and communities may wish to consider when evaluating funding for grade crossing warning system upgrades.

5.5.4. Model Validation

The models in Section 5.4 were validated using data from 2011 and 2012, the two most recent years for which data were available. The purpose of this validation was to test the models' predictive ability against unseen data, as well as to ensure that their predictive ability does not decline over time.

I considered trying to validate the models by comparing the observed and expected number of derailments using yearly data as the observed values for number of derailments and the model predictions as the expected. However, I determined this is not feasible. Developing an expected value using the models would require selection of a cut point or threshold value, and as discussed in Section 5.5.3, selection of this value is essentially a policy decision based on tradeoffs between false positives and false negatives. To ensure identification of most derailment incidents, the model has a high false negative rate (compared to its false positive rate). Therefore the number of "expected" derailments is always going to be larger than the observed number.

Instead, I evaluated the model's performance by testing it against new data. The validation was performed using the SCORE statement in SAS that "scores" a new dataset using a model trained on another dataset. The characteristics of the validation datasets are shown in Table 5.10.

Table 5.10: Summary of Datasets Used For Validation

Dataset	Number of Derailments	Number of Non-Derailments	Total Number of Events
AV	26	2,295	2,321
AV-VST	7	364	371
AV-TSV	19	1,931	1,950
BV	26	206	232

5.5.4.1. Unified Model

For the unified model presented in Section 5.4.1, the validation was performed using dataset AV, which represents all of the highway-rail grade crossing collisions that occurred in 2011 and 2012. For dataset AV, tested with the unified model, the area under the ROC curve (Figure 5.5) was found to be 0.7616, which shows the model has an acceptable ability to discriminate between derailment and non-derailment events regardless of the chosen threshold (Section 5.5.3).

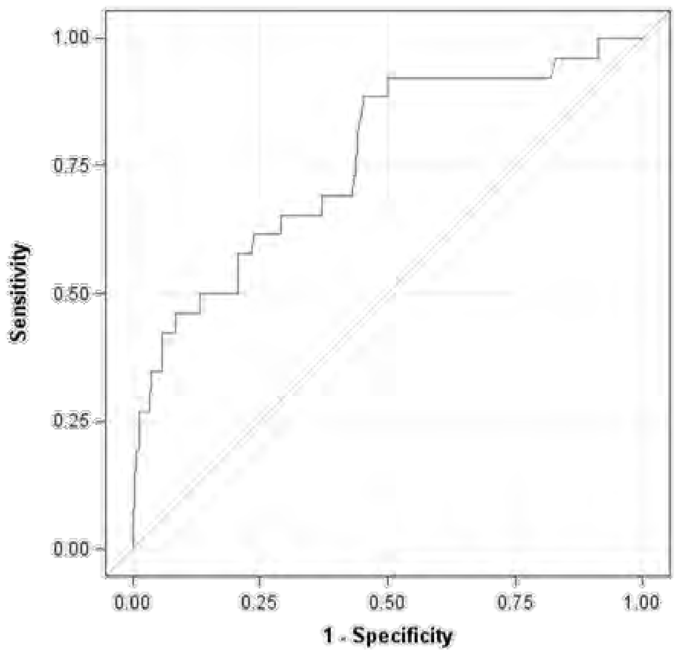


Figure 5.5: ROC Curve for dataset AV using the unified derailment model.

The SAS software does not provide an HL test value for scored data; however, it does provide a Brier score. The Brier score of a dataset is a measure of goodness-of-fit; it represents the difference between the predicted probability and the observed response of a data point (Chapter 3). For the validation dataset fitted with the unified model, the Brier score is 0.0722, which indicates a good fit of the model.

5.5.4.2. Split Model

For the split model presented in Section 5.4.2, the validation was performed using datasets AV-VST and AV-TSV. AV-VST represents all the highway-rail grade crossing incidents in 2011 and 2012 where the highway user struck the train. AV-TSV represents grade crossing incidents where the train struck the highway user. For dataset AV-VST, tested with the split model for incidents where the highway user struck the train, the area under the ROC curve (Figure 5.6) was found to be 0.8562, which shows the model has good ability to discriminate between derailment and non-derailment events. For this validation dataset, the Brier score is 0.0501, which indicates a good fit of the model.

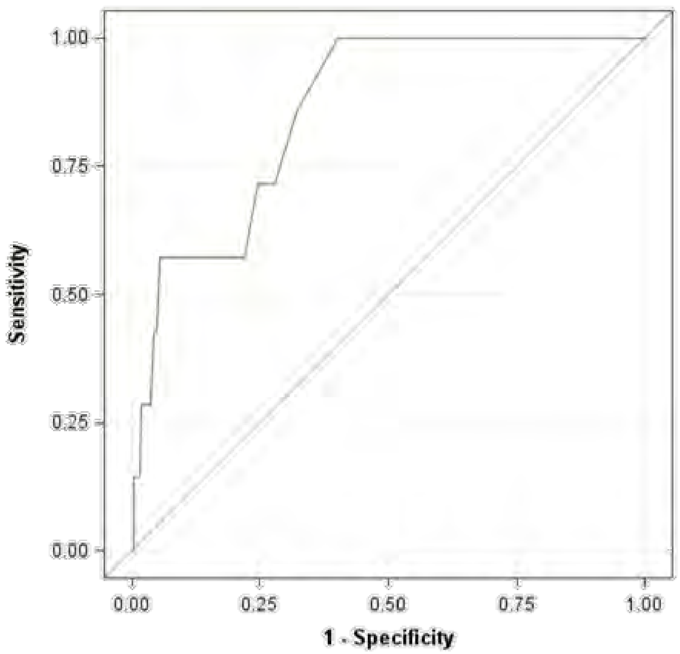


Figure 5.6: ROC Curve for validation dataset AV-VST.

For dataset AV-TSV, tested with the split model for incidents where the train struck the highway user, the area under the ROC curve was found to be 0.8384, which shows the model has good ability to discriminate between derailment and non-derailment events (Figure 5.7). For this validation dataset, the Brier score is 0.1084, which shows acceptable fit of the model.

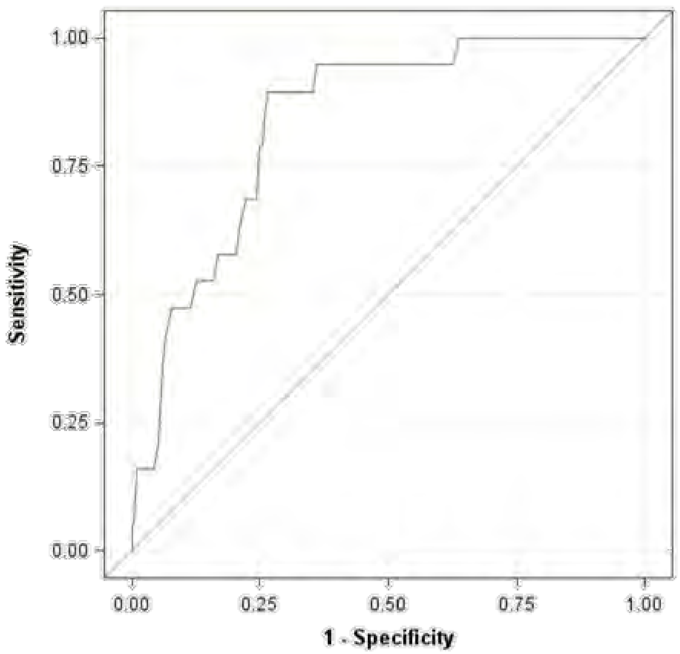


Figure 5.7: ROC Curve for validation dataset AV-TSV.

5.5.5. Model Limitations

As with any regression model, these findings are limited by the quantity and quality of data available. Derailments due to grade crossing incidents are uncommon events. Development of a reasonably-sized dataset of incidents required use of 20 years of data during which there were 312 verified derailment events. A possible concern is that factors pertinent to this investigation may have changed over the period encompassed by the dataset. Grade crossing incident rates have declined by approximately 60% since 1991. This is due in large part to programs such as Operation Lifesaver and changes in driver behavior, as well as crossing closure or reconstruction and FRA-mandated reflectorization of rolling stock (Mok and Savage, 2005; 68 FR 215, 2003; Chaudhary et al., 2011). If the data are not homogenous over the 20-year period of study, it could result in unexplained variance. While the number of REA-reportable grade crossing incidents has decreased, the percentage of grade crossing incidents that result in derailment has remained approximately constant (Figure 5.8).

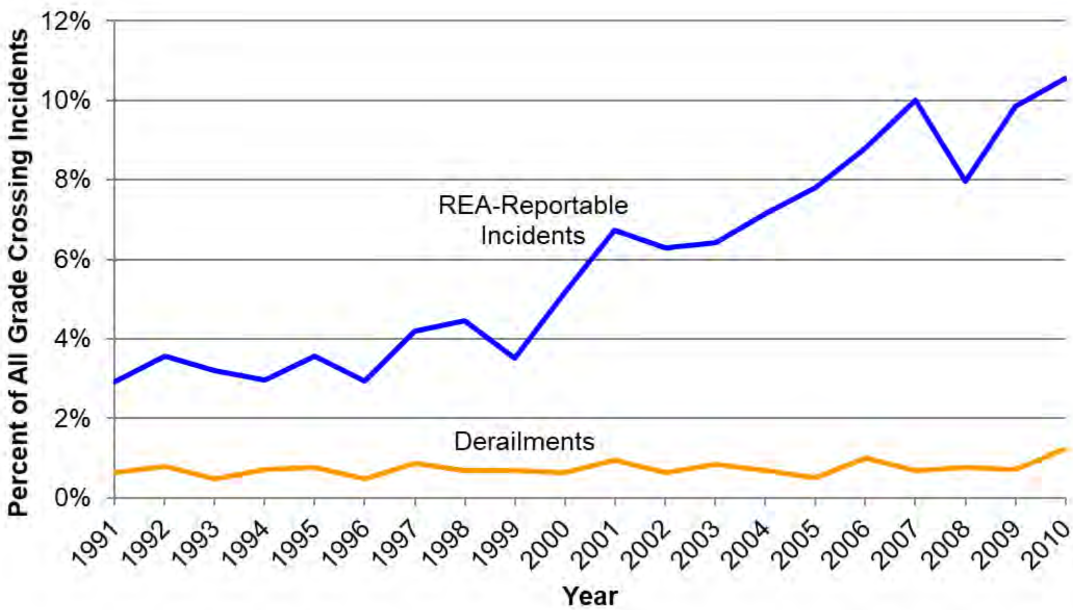


Figure 5.8: Percent of grade crossing incidents that were REA-reportable collisions or that were derailments, 1991-2010.

Interestingly, the percent of REA-reportable grade crossing incidents has increased from less than 3% to just over 10% in the same time period. While the reason for this increase is unclear, one hypothesis may be that the reporting threshold for the REA is increasing at a lesser rate than the actual increase in incident cost to the railroads. This would mean that a greater proportion of incidents are being reported every year, even though those incidents are not necessarily more severe. The reporting threshold is determined each year based on an equation developed by the Federal Railroad Administration (FRA, 2014) with the threshold values periodically adjusted for inflation (Table 5.11).

Alternatively, this change could be due to the increase in the number of crossings that have active warning devices (Mok and Savage 2005). If a collision involves damage to an active warning system, the cost of an incident is more likely to exceed the reporting threshold because these systems are considerably more expensive than passive warning systems subject to damage in a derailment. It is outside the scope of this dissertation to assess this hypothesis, but future researchers could investigate it using data contained in the REA and HRA databases. These break down the cost of incidents into multiple categories, including "equipment damage" (damage to the train consist) and "track damage" (damage to the track, signals, right-of-way and structures). If the cost of track damage is increasing at a higher rate than the cost of equipment damage, it would be consistent with this hypothesis.

Table 5.11: REA Reporting Thresholds, 1991-2013 (FRA, 2014)

Year	Threshold	Year	Threshold
1991	$ 6,300	2003	$ 6,700
1992	$ 6,300	2004	$ 6,700
1993	$ 6,300	2005	$ 6,700
1994	$ 6,300	2006	$ 7,700
1995	$ 6,300	2007	$ 8,200
1996	$ 6,300	2008	$ 8,900
1997	$ 6,500	2009	$ 8,900
1998	$ 6,600	2010	$ 9,200
1999	$ 6,600	2011	$ 9,400
2000	$ 6,600	2012	$ 9,500
2001	$ 6,600	2013	$ 9,900
2002	$ 6,700		

Overall, the quality of the data is good. There are some errors and inconsistencies between the REA and HRA databases, but in general it is a relatively simple matter to identify and correct these errors. There are sufficient data so incomplete or incorrect records can be dropped if they cannot be corrected. The greatest challenge is to identify events resulting in a derailment. It was assumed that the majority of derailment events are recorded in the REA database. This claim can be partially verified for incidents occurring between 1997 and 2010. Beginning in 1997, a narrative field was added to the HRA database. In most instances of derailment, the person filing the report used this field to mention that a derailment had occurred. Therefore, even though there is no derailment variable in the HRA database, derailment information could sometimes be extracted. Through this method, nine grade crossing derailments were found in the HRA database that were not reported in the REA database, probably because they did not exceed the REA damage threshold. There were 312 grade-crossing caused derailments during the same interval, indicating that the majority of derailment events have been captured in this analysis. However, it would be useful if the HRA database started tracking derailments as well.

5.6. CONCLUSIONS

This chapter described development of a model to predict derailment rates at highway rail grade crossings involving freight trains using logistic regression modeling. Results show that four of the nine analyzed factors are important to the model: incident type, highway vehicle type, highway vehicle speed and train speed. Two separate models were developed: a simpler one using a unified dataset, and a more complex but more accurate model using a split dataset.

CHAPTER 6: JOINT MODELING OF PASSENGER AND FREIGHT TRAIN DERAILMENT LIKELIHOOD

6.1. INTRODUCTION

This chapter focuses on development of a model that estimates the probability of a grade crossing collision resulting in either a passenger or freight train derailment, based on a variety of characteristics of the collision. The previous chapter developed statistical models for freight trains, but the goal was to expand the model to work with passenger trains as well. While most crossings in the US do not have freight trains, a large number have a combination of freight and passenger trains, so being able to model the derailment likelihood of both will better represent risk at the crossing, and enhance the utility of the model in other ways as well.

6.2. JOINT FREIGHT-PASSENGER MODEL

Due to the relative lack of data on grade-crossing-caused derailments of passenger trains, it was difficult to develop an independent model capable of accurately predicting this type of incident. Only 62 derailments involving passenger trains have occurred in the past 20 years, meaning that there is insufficient information to model the interaction of four (or more) factors. Therefore, I pursued an alternate approach in which freight and passenger data were combined into one model.

The original motivation for separating freight and passenger data was twofold. Passenger and freight trains differ in their physical and operational characteristics and it was believed that limiting variation in the dataset would improve the accuracy of exploratory analysis and modeling. Second, I had already found that freight and passenger trains have different likelihood of derailment in grade crossing incidents. While 0.7% of grade crossing collisions involving freight trains result in derailment, 1.2% of passenger train collisions result in derailment. Therefore, passenger trains are almost twice as likely to derail as freight trains. This difference in rate implied an inherent difference in derailment likelihood for passenger and freight trains.

Passenger trains differ from freight trains in three key ways: they generally operate at higher speeds, are considerably shorter in length, and use lighter-weight rolling stock. The freight analysis showed that derailment likelihood was dependent on three physical factors: the size (or weight) and speed of the highway vehicle, and the speed of the train. By adding consideration of rolling stock weight and train length – the remaining factors from the physical description of derailments from Chapter 1 – to the model, it could be possible to combine the freight and passenger datasets.

Combining freight and passenger data into one model has multiple benefits. First, by providing more data, it increases the power of the statistical models. Second, considering the impact of rail equipment type and weight improves understanding of the effect these factors have on derailment occurrence. Last, a joint freight-passenger model is more useful to practitioners, since many crossings in the U.S. have both freight and passenger traffic. A combined model will more accurately portray risk at these crossings.

6.3. METHODOLOGY

6.3.1. Database Development and Candidate Variables

Three FRA databases were used in the analysis described in this chapter: the Rail Equipment Accident/Incident database, the Highway Rail Accident (HRA) database, and the Grade Crossing Inventory (GCI). Data for all U.S. mainline railroads (both freight and passenger) during the 20-year period 1991 through 2010 were used. Note that the number of events shown in Table 6.1 is less than the numbers used in Chapter 5. This is because inclusion of the train length and railroad equipment type factors required additional fields to have information in the HRA database. Due to a combination of missing data and the elimination of unclear data, there are approximately 10,000 fewer records in this analysis than the previous analysis. However, this leaves over 30,000 records for analysis, and there is no reason to suspect there is bias to the removed records.

Table 6.1: Summary of Data Used in Model Development

Dataset	Number of Derailments	Number of Non-Derailments	Total Number of Events
VST – Passenger	6	393	399
VST – Freight	91	6,550	6,641
TSV-S – Passenger	17	1,273	1,290
TSV-S – Freight	43	9,975	10,018
TSV-M – Passenger	16	1,400	1,416
TSV-M – Freight	98	13,513	13,611

The goal of this chapter is to develop a statistical model to predict the probability of a highway-rail grade crossing collision resulting in a derailment, regardless of whether the train is a passenger or freight train. The model has a binary response variable: either there is a derailment or there is not. A variety of input variables were considered that described physical characteristics of the grade crossing as well as accident-related factors. The input variables were binary, categorical, or continuous in nature and are summarized in Table 6.2.

Table 6.2: Definition of Model Variables

Variable Name	Definition	Variable Type	Range of Values
VS	Highway vehicle estimated speed in mph	Continuous	Range*: 0-79 mph Average*: 9.53 mph Standard Deviation*: 11.94
TS	Train speed in mph	Continuous	Range*: 0-106 mph Average*: 51.16 mph Standard Deviation*: 23.02
LV	Was a large highway vehicle involved?	Binary	N if no; Y if yes
IT	Incident type (Did train strike highway user?)	Categorical	VST if highway user struck train; TSV-S if train struck stationary highway user; TSV-M if train struck moving highway user
EC	Equipment class	Categorical	FC if freight railcar; FL if freight locomotive; PC if passenger railcar; PL if passenger locomotive
TL	Train length	Continuous	Range*: 1-217 Average*: 54.85 Standard Deviation*: 39.46
TC	FRA track class	Categorical	0-9 (0 represents X)
WS	Crossing warning interconnected with highway signals	Categorical	1 if yes; 2 if no; 3 if unknown
VIEW	Was the driver's view of the track obstructed?	Binary	N if not obstructed; Y if obstructed
PUBLIC	Did the collision occur at a public crossing?	Binary	Y if public; N if private
XTYPE	Type of warning device at crossing	Categorical	1: gates 2: active (excl. gates) 3: passive 4: other 5: none

* Note: Statistics are for all data (not divided by freight/passenger or incident type).

Incident type, train speed, highway vehicle speed and vehicle type – the factors in the freight model from Chapter 5 – were also expected to be factors in the joint passenger-freight model. The effect of train mass was added and represented as two variables – train length (TL) and equipment class (EC). Track class, crossing visibility, type of warning device and accessibility of crossing (public or private) were considered in the model, as they were in the freight-only model. Second-order interaction and polynomial effects were considered for the continuous variables, as well as interactions for the categorical variables.

While the quality of the HRA and REA databases is excellent overall, some records were either internally inconsistent or had missing data. Additionally, some fields were re-coded to make them easier to use in the model (see data cleanup procedure in Appendix A).

During the analysis process the inclusion of "auto trains" in the FRA databases became apparent and led to some confounding results. Auto trains are a special Amtrak service that only operates in the eastern United States. These trains are unique in that they are a mix of passenger and freight equipment. They are substantially different from other passenger trains in several respects, notably their much longer length and auto-carrying railcars (Figure 6.1).

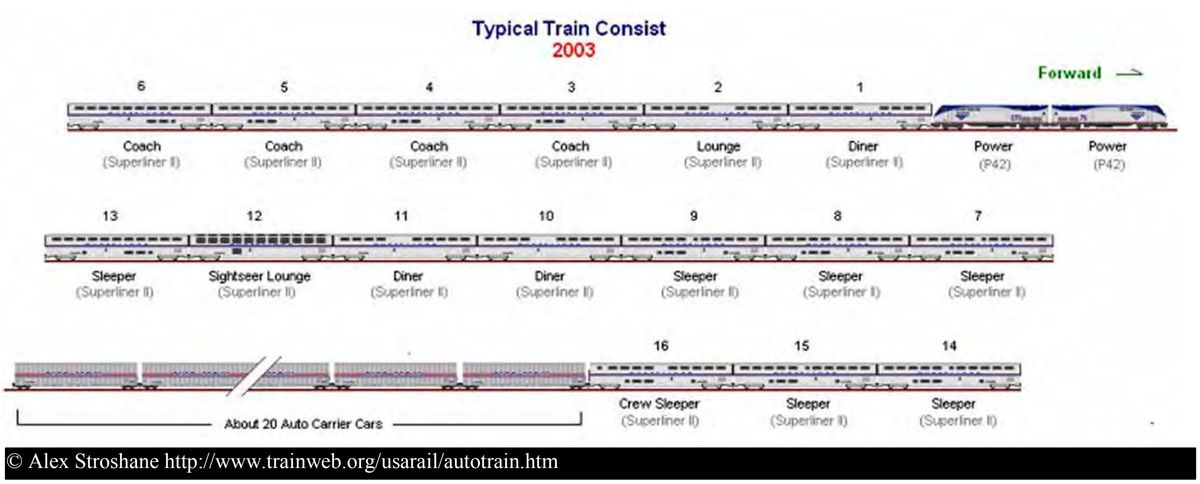


Figure 6.1: Example of Amtrak Auto Train.

Therefore, they cannot be simply classified as either passenger or freight. Because of their mixture of characteristics, I excluded them from the analysis by removing any records for passenger trains with more than 31 rail vehicles in the consist. This threshold is an assumption based on information in the narrative fields of the REA database, indicating that some trains were auto trains. In regular service today, it is unlikely for a passenger train to have more than 20 rail vehicles in the consist. Trains from the HRA and REA databases with between 20 and 30 rail vehicles were mostly part of Amtrak's "Mail and Express" service in the late 1990s and early 2000s. The service has since been discontinued, but these records were left in the dataset because the mail cars had similar characteristics to other passenger cars.

6.3.2. Statistical Modeling Technique

The model was developed using the LOGISTIC procedure in the Statistical Analysis Software (SAS) program, as described in Chapter 3. Based on the lessons learned during development of the freight models, I began with two models based on incident type – one each for VST and TSV incidents. After further consideration and analysis, I subdivided the TSV incidents into two categories based on whether the highway vehicle was moving or not. These two categories are TSV-S (train strikes stopped vehicle) and TSV-M (train strikes moving vehicle). About 43% of TSV incidents involve vehicles that are stopped on the crossing. The impact forces associated with a stopped vehicle are different than those involving a moving vehicle. It seemed plausible that this might mask the true effect of highway vehicle speed when conducting statistical regression. The same problem did not exist for VST incidents, since by definition the highway vehicle is moving and very few (less than five out of the whole database) trains involved in VST incidents were stopped at the time of collision.

As with the freight models, I adjusted for the underrepresentation of derailment events in the overall dataset using the “rare events logistic regression” (RELR) technique (discussed in Chapter 3).

6.4. RESULTS

6.4.1. VST Incidents

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

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

This “best model” is:

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

$$x_{VST} = -2.0204 + 0.0607\,VS + \begin{cases} 0, & LV = Y \\ -1.5458, & LV = N \end{cases} + \begin{cases} 1.8213, & EC = PC \\ 0.0648, & EC = FC \\ 0, & EC = PL \\ -1.3087, & EC = FL \end{cases} \tag{1}$$

(s.e.) (0.4179) (0.0099) (0.2081) (0.7159; 0.3589; 0.3737)[1]

where VS, LV, and EC are as described in Table 6.2.

This model provided the best fit to the data, with a Hosmer-Lemeshow (HL) goodness-of-fit test result of 0.7222. Values closer to 1 indicate good model fit, and values closer to 0 indicate poor fit. This model also has the ability to discriminate between derailment and non-derailment events, as measured by the ROC curve (Figure 6.2). Generally, a model is considered to provide good discrimination if the ROC value is greater than 0.8. The area under the ROC curve for this result is 0.9011.

[1] These standard errors correspond to the difference between each of the EC categories and the "baseline category", in this case "passenger locomotive" (PL).

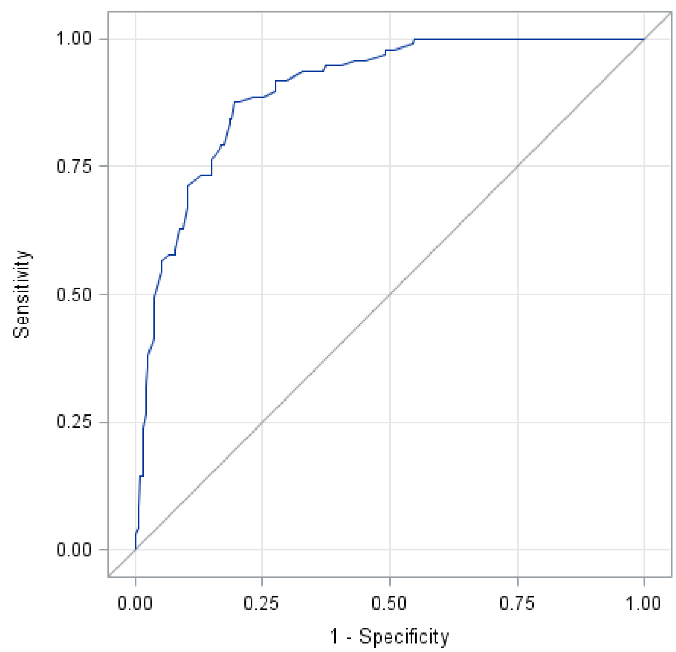


Figure 6.2: ROC Curve for retrospective model. Area under the curve is equal to 0.9011.

Additional performance statistics for this model are given in Table 6.3. For these values, the threshold value for predicting a derailment was a p value of 0.3. If the calculated value of p for a data point was greater than 0.3, it was classified as predicting a derailment, and if it was less than 0.3, it was classified as predicting no derailment.

Table 6.3: Performance Statistics for Retrospective Model

Statistic	Cases
Percent Correct	81.8
Sensitivity	86.6
Specificity	79.5

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

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

The intercept term is then modified to account for $p_{all\ VST}$ using the log-odds operator (*6*).

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

$$b_{VST} = -2.0204 + \ln\left(\frac{0.0138}{1 - 0.0138}\right) = -6.2912$$

Using the modified intercept term adjusts the probabilities predicted by the model to reflect the actual observed rate of derailments. For all VST incidents, the following model will be used:

$$p_{all\,VST} = \frac{1}{e^{-x_{all\,VST}} + 1}$$

$$x_{all\,VST} = -6.2912 + 0.0607\,VS + \begin{cases} 0, & LV = Y \\ -1.5458, & LV = N \end{cases} + \begin{cases} 1.8213, & EC = PC \\ 0.0648, & EC = FC \\ 0, & EC = PL \\ -1.3087, & EC = FL \end{cases} \tag{2}$$

(s.e.) (2.1921) (0.0197) (0.3522) (0.0195)

where VS, LV, and EC are as described in Table 6.2.

An ROC curve was generated by analyzing the total population dataset with equation $p_{all\ VST}$ (Figure 6.3). The area under the ROC curve was 0.9056, which indicates good discrimination. Additionally, model performance was quantified using the Brier score. This model had a Brier score of 0.0809; Brier scores closer to zero indicate better fit.

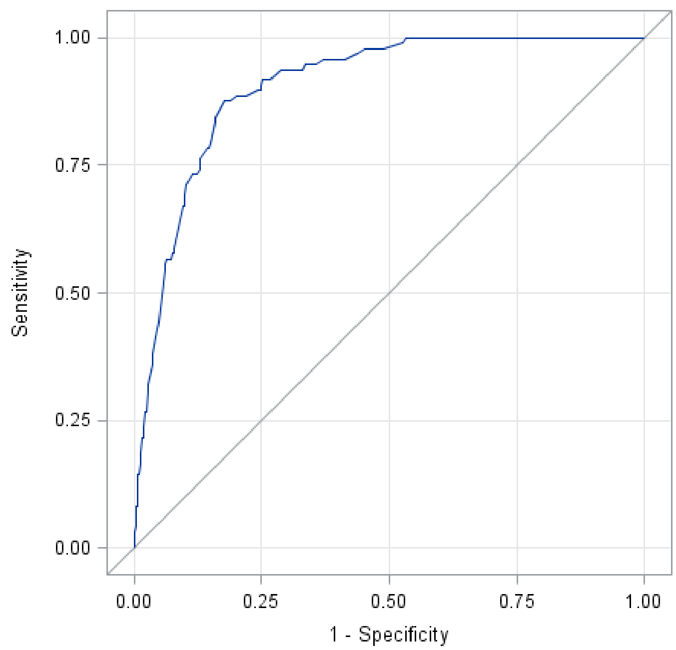


Figure 6.3: ROC Curve for dataset VST. Area under the ROC curve is equal to 0.9056.

In addition to these traditional techniques, the model was tested to see how it performed at ranking incidents by derailment likelihood, and whether this ranking corresponded to whether a derailment actually occurred. This technique has the advantage of being independent of the selected threshold value. To do this, all VST incidents in the HRA database were ranked by their $p_{all\ VST}$ value as calculated by the model, from least likely to most likely to derail. The dataset was divided into quintiles and the number of derailments in each quintile were counted (Table 6.4).

Table 6.4: Performance of VST Model Based on Ranking

Quintile	Assigned Rank	Actual Derailments	Percent of Actual Derailments in Quintile
1 (least likely to derail)	0 – 1,408	0	0%
2	1,409 – 2,816	1	1.03%
3	2,817 – 4,224	7	7.22%
4	4,225 – 5,632	12	12.37%
5 (most likely to derail)	5,633 – 7,040	77	79.38%

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

6.4.2. TSV-S Incidents

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

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

This "best model" is:

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

$$x_{TSV-S} = \quad -5.2729 \; + 0.0893\, TS \; + 0.0362\, TL \; - 0.00075\, TS \times TL \; + \begin{cases} 0, & LV = Y \\ -1.5458, & LV = N \end{cases} \quad (3)$$

(s.e.) (1.2119) (0.0231) (0.0139) (0.0003) (0.2548)

where TS, TL, and LV are described in Table 6.2.

This model had an HL test result of 0.8535 and an area under the ROC curve of 0.8688 (Figure 6.4). Additional performance statistics for this model are given in Table 6.5 for a p-value of 0.3.

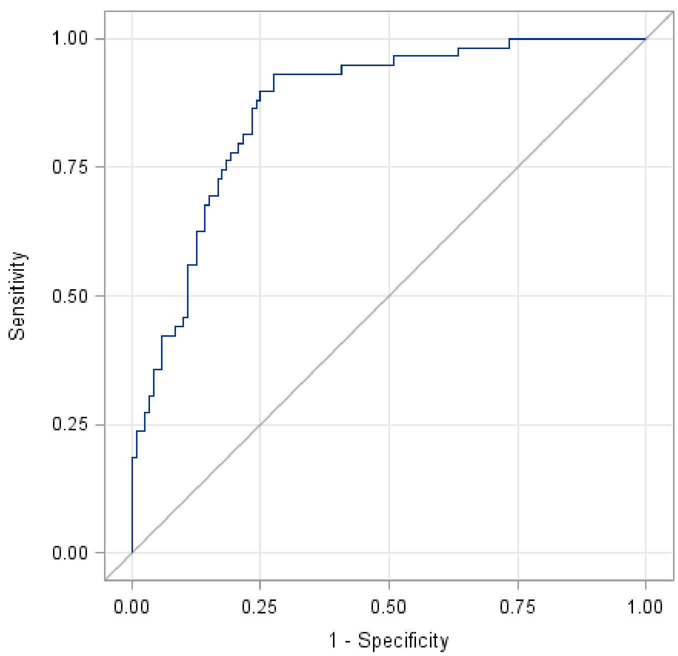


Figure 6.4: ROC Curve for TSV-S retrospective model. Area under the curve is equal to 0.8688.

Table 6.5: Performance Statistics for Retrospective Model

Statistic	Cases
Percent Correct	78.2
Sensitivity	88.1
Specificity	73.3

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

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

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

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

$$b_{TSV-S} = -5.2729 + \ln\left(\frac{0.0053}{1 - 0.0053}\right) = -10.5065$$

For all TSV-S incidents, the following model will be used:

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

$$x_{all\ TSV-S} = -10.5065 + 0.0893\,TS + 0.0362\,TL - 0.00075\,TS \times TL + \begin{cases} 0, & LV = Y \\ -1.5458, & LV = N \end{cases} \quad (4)$$

where TS, TL, and LV are as described in Table 6.2.

The area under the ROC curve for the total population dataset was 0.8790, which is considered good discrimination (Figure 6.5). The model has a Brier score of 0.0892 indicating good fit.

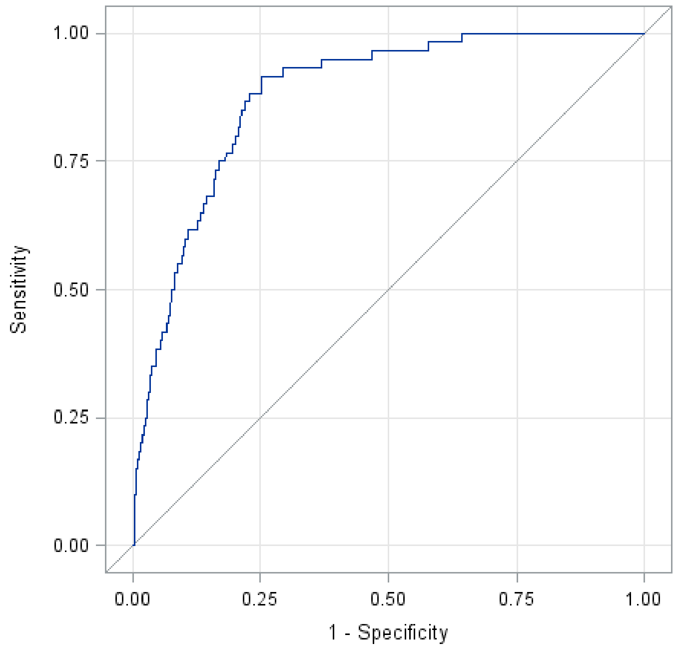


Figure 6.5 ROC Curve for dataset TSV-S. Area under the ROC curve is equal to 0.8790.

I ranked all TSV-S incidents by their $p_{all\ TSV\text{-}S}$ value calculated by the model, from least likely to most likely to derail. I then divided the dataset into quintiles and counted how many actual derailments occurred in each quintile (Table 6.6).

Table 6.6: Performance of TSV-S Model Based on Ranking

Quintile	Assigned Rank	Actual Derailments	Percent of Actual Derailments in Quintile
1 (least likely to derail)	0 – 2,261	0	0%
2	2,262 – 4,522	1	1.67%
3	4,523 – 6,783	2	3.33%
4	6,784 – 9,044	10	16.67%
5 (most likely to derail)	9,045 – 11,305	47	78.33%

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

6.4.3. TSV-M Incidents

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

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

This "best model" is:

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

$$x_{TSVNZ} = -3.2144 + 0.0243\,VS + 0.0233\,TS + \begin{cases} 0, & LV = Y \\ -2.2628, & LV = N \end{cases} \tag{5}$$

(s.e.) (0.5574) (0.0117) (0.0086) (0.3666)

where VS, TS, and LV are described in Table 6.2.

This model provided the best fit to the data, with an HL goodness-of-fit test result of 0.9152 and an area under the ROC curve of 0.8438 (Figure 6.6). Additional performance statistics for this model are given in Table 6.7.

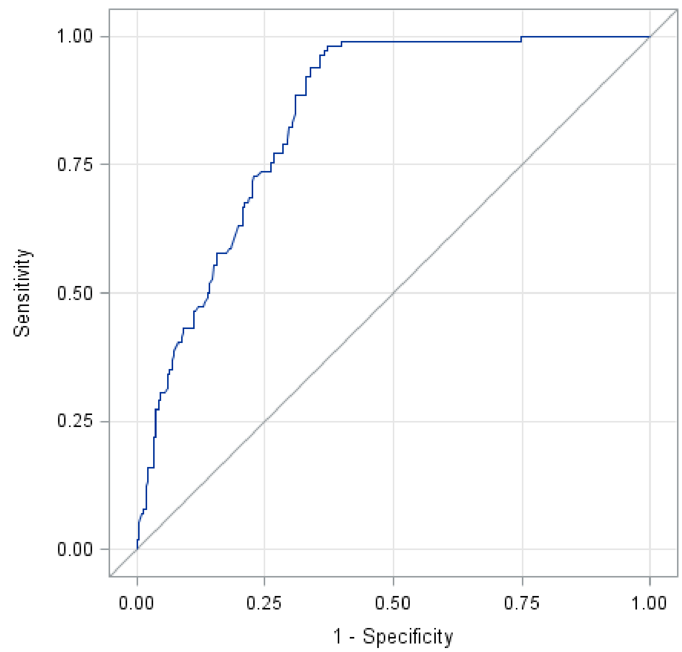


Figure 6.6: ROC Curve for TSV-M retrospective model. Area under the curve is equal to 0.8438.

Table 6.7: Performance Statistics for Retrospective Model

Statistic	Cases
Percent Correct	74.1
Sensitivity	98.2
Specificity	61.5

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

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

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

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

$$b_{TSV-M} = -3.2144 + \ln\left(\frac{0.0076}{1 - 0.0076}\right) = -8.0882$$

For all TSV-M incidents, the following model will be used:

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

$$x_{TSV-M} = \quad -8.0882 \quad + 0.0243\, VS + 0.0233\, TS \quad + \begin{cases} 0, & LV = Y \\ -2.2628, & LV = N \end{cases} \tag{6}$$

where VS, TS, and LV are as described in Table 6.2.

The area under the ROC curve created by evaluating all records with $p_{all\ TSV\text{-}M}$ was 0.8625, which indicates good discrimination (Figure 6.7). The model has a Brier score of 0.1038.

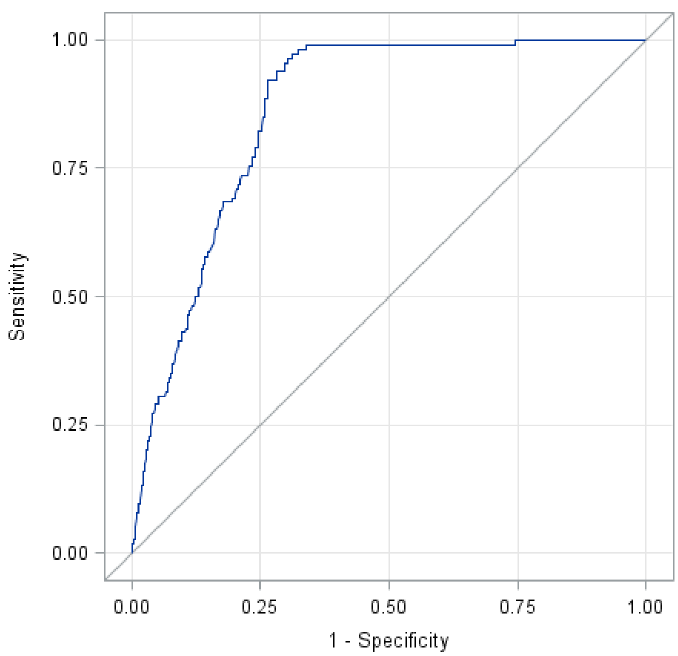


Figure 6.7: ROC Curve for TSV-M. Area under the ROC curve is equal to 0.8625.

All TSV-M incidents in the model development dataset were ranked by their $p_{all\ TSV\text{-}M}$ value calculated by the model, from least likely to most likely to derail. The dataset was divided into quintiles and the actual derailments occurring in each quintile were counted (Table 6.8).

Table 6.8: Performance of TSV-M Model Based on Ranking

Quintile	Assigned Rank	Actual Derailments	Percent of Actual Derailments in Quintile
1 (least likely to derail)	0 – 3,006	0	0%
2	3,007 – 6,012	1	0.88%
3	6,013 – 9,018	0	0%
4	9,019 – 12,024	0	0%
5 (most likely to derail)	12,025 – 15,027	113	99.12%

Since nearly 100% of actual derailment incidents were ranked in the 5th quintile, the model did a good job identifying derailment incidents.

6.4.4. Model Validation

To verify that the models developed based on data from 1991-2010 were valid for incidents outside the study period, data from incidents between 2011-2015 were tested to see where derailment incidents would end up in the ranking. The same technique used to develop Tables 6.4, 6.6 and 6.8 was used to analyze the model's performance with the recent data. For each of the three incident types, at least 73% of the derailments that occurred between 2011 and 2015 ranked in the fifth quintile (incidents with the highest conditional probability of derailment). These results showed that the model performs as well for more recent incidents as it did for incidents in the development dataset (Table 6.9).

Table 6.9: Performance Metrics for Validation Dataset (2011-2015)

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

6.5. DISCUSSION

6.5.1. Interpretation of Model Terms

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

6.5.1.1. VST Incidents

The first term in the VST model (Equation 2), $0.0607\ VS$, indicates that the speed of the vehicle at collision affects derailment likelihood. As vehicle speed increases, the probability of derailment also increases, consistent with the univariate analysis (Chapter 4) and the findings for freight incidents (Chapter 5). The standard error is less than the value of the coefficient, so while the exact value of the coefficient may vary, the effect is always positive, meaning that increased vehicle speed increases derailment likelihood.

The second term in the model, $\begin{cases} 0, & LV = Y \\ -1.5458, & LV = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle such as a car, motorcycle or pickup-truck (*LGVEH* = *N*) then this term assumes a value of -1.5458; if the highway user is a large vehicle such as a tractor-semi-trailer or a straight truck (*LGVEH* = *Y*) then the term disappears. This means that, all else equal, a collision involving a large highway vehicle is more likely to result in a derailment. This result is consistent with the findings described in Chapters 4 and 5.

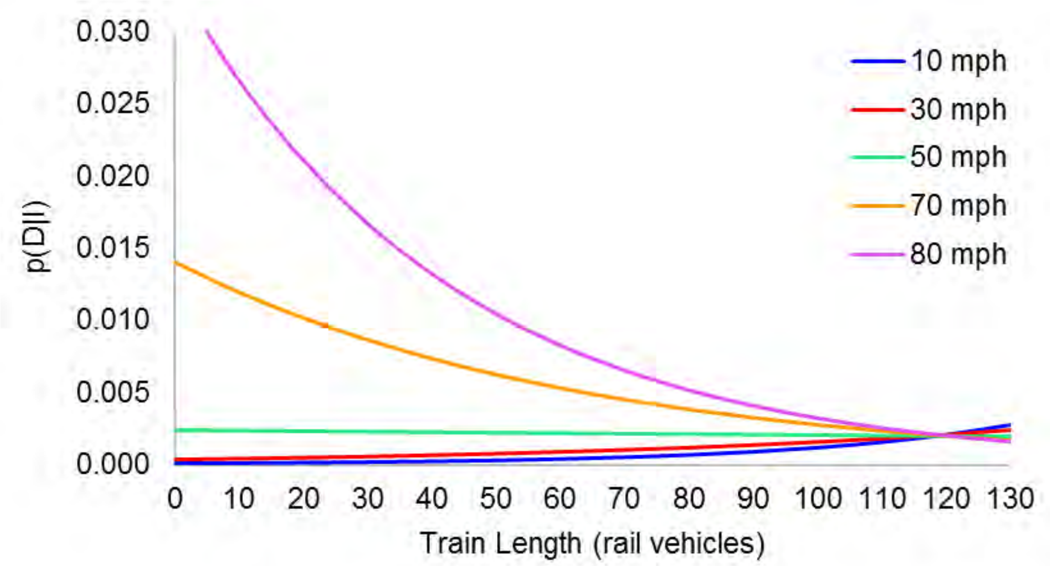
The third and final term in the model, $\begin{cases} 1.8213, & EC = PC \\ 0.0648, & EC = FC \\ 0, & EC = PL \\ -1.3087, & EC = FL \end{cases}$, shows the effect of equipment class on derailment likelihood. As the coefficients become more positive, derailment likelihood increases. Incidents involving passenger railcars (PC) are more likely to result in derailment than those involving freight railcars (FC), which in turn are more likely to result in derailment than those involving passenger locomotives (PL), which in turn are more likely to result in derailment than those involving freight locomotives (FL). This trend is consistent with the hypothesis that lighter rail equipment is more likely to derail than heavier rail equipment. It should be noted that the confidence intervals of the estimates for freight railcars and passenger locomotives overlap, meaning it is statistically uncertain if there is a difference between these two equipment classes. This overlap was observed in all four of the candidate models, suggesting it is not an artifact of this particular dataset. The overlap may be explained by the wide range in freight car weight. A loaded freight railcar can weigh five times as much as an empty one. When unloaded, the average freight railcar is lighter than the average passenger locomotive, but the opposite is true for loaded freight railcars. Unfortunately, the HRA database does not track whether the railcar was loaded or empty (unlike the REA, which does provide this information); therefore it was not possible to distinguish between loaded and empty freight cars.

6.5.1.2. TSV-S Incidents

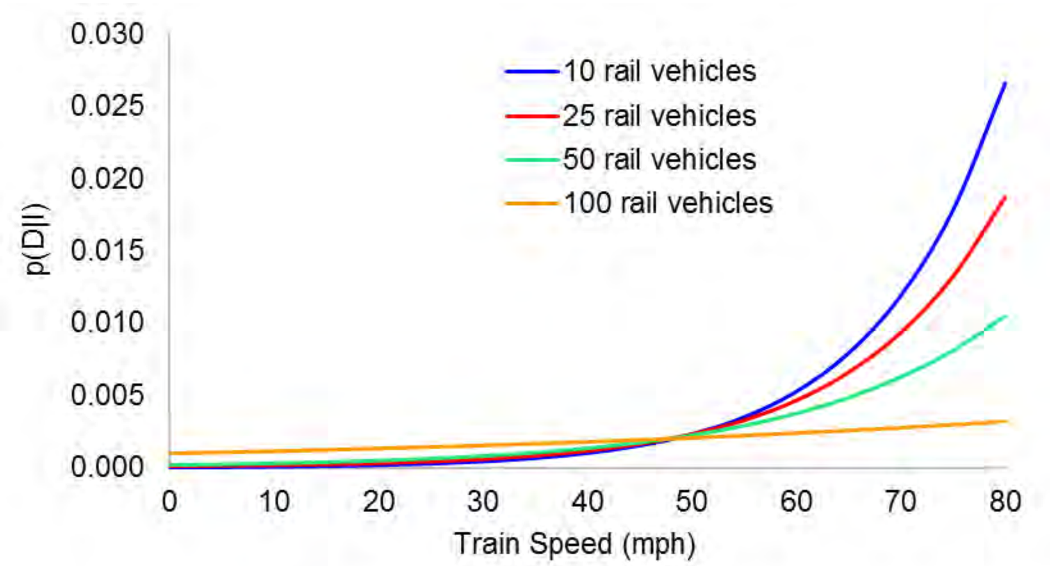
The first term in the TSV-S model (Equation 4), $0.0893\,TS$, indicates that the speed of the train at collision affects derailment likelihood. As train speed increases, the probability of derailment also increases, consistent with the univariate analysis (Chapter 4) and the findings for freight incidents (Chapter 5). The standard error is less than the value of the coefficient, so while the exact value of the coefficient can vary, the effect is always positive; increased train speed increases derailment likelihood.

The second term in the model, $0.0362\,TL$, indicates that there is a relationship between train length and derailment likelihood. As the length of the train increases, so does derailment likelihood. Since the standard error is less than the coefficient, the effect is positive.

The third term in the model, $-0.00075\,TS{\times}TL$, is an interaction effect between train speed and train length. This indicates that at higher train speeds, the likelihood of derailment decreases with longer train length; or, alternatively, for longer trains, the likelihood of derailment decreases with increased speed. To fully understand the effect of train speed and train length, it is necessary to consider the first, second and third terms collectively. The effect of these two variables on derailment likelihood was investigated in two ways. First, derailment likelihood was plotted as a function of train length for different train speeds (all other factors were held constant) (Figure 6.8 (a)). Second, it was plotted as a function of train speed for different train lengths (Figure 6.8(b)).



(a)



(b)

Figure 6.8: Illustration of the interaction effect of first, second and third terms for TSV-S model, with (a) derailment likelihood as a function of train length, and (b) derailment likelihood as a function of train speed.

The final term in the model, $\begin{cases} 0, & LV = Y \\ -1.5733, & LV = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle such as a car or pickup-truck (*LGVEH* = *N*) then this term assumes a value of -1.5733; if the highway user is a large vehicle such as a tractor-semi-trailer or a straight truck (*LGVEH* = *Y*) then the term disappears. *Ceteris paribus*, a collision where the highway user is a large vehicle is more likely to result in a derailment, consistent with the results presented in Chapters 4 and 5.

6.5.1.3. TSV-M Incidents

The first term in the TSV-M model (Equation 6), $0.0243\,VS$, indicates that the speed of the vehicle at collision affects derailment likelihood. As vehicle speed increases, the probability of derailment also increases, consistent with the univariate analysis (Chapter 4) and the findings for freight incidents (Chapter 5). The standard error is less than the value of the coefficient, so while the exact value of the coefficient can vary, the effect is always positive, so increased vehicle speed increases derailment likelihood.

The second term in the model, $0.0233\ \mathrm{TS}$, indicates that the speed of the train at collision affects derailment likelihood. As train speed increases, the probability of derailment also increases, consistent with Chapters 4 and 5. The standard error is less than the value of the coefficient, so although the exact value of the coefficient can vary, the effect is positive, so increased train speed increases derailment likelihood.

The final term in the model, $\begin{cases} 0, & LV = Y \\ -2.2628, & LV = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle (*LGVEH* = *N*) then this term assumes a value of -2.2628; if the highway user is a large vehicle (*LGVEH* = *Y*) then the term disappears. This means that, all else equal, a collision where the highway user is a large vehicle is more likely to result in a derailment, consistent with Chapters 4 and 5.

6.5.2. Model Limitations

As with any model, these findings are limited by the quantity and quality of data available. Derailments due to grade crossing collisions are uncommon events, and despite their higher rate of occurrence given an incident, passenger train derailments are especially uncommon. Development of a reasonably-sized dataset of accidents required use of 20 years of data during which there were 56 verified derailments involving passenger trains.

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

Clean-up for the passenger dataset differed from the freight data in one major aspect – the inclusion of "auto-trains", which contain a combination of passenger and freight railcars. These are unusual trains that only run on one rail line in the US, and are much longer than typical passenger trains. Due to their mixture of freight and passenger train characteristics, they were removed from the dataset to eliminate the additional variance they would introduce. Therefore, this model will not necessarily represent derailment likelihood at crossings with auto-trains.

6.6. CONCLUSIONS

This chapter explored the development of a set of models to predict derailment rates for both freight and passenger trains at highway-rail grade crossings using logistic regression analysis. Three regression models were ultimately developed based on incident type – one each for incidents where the vehicle strikes the train, incidents where a train strikes a stopped vehicle, and incidents where a train strikes a moving vehicle. Results show that, other than incident type, five factors are important to derailment prediction: highway vehicle type, highway vehicle speed, train length, rail equipment type, and train speed. The key factors varied for each of the three regression models in ways that are consistent with expectations given the physical forces for each incident type.

The next chapter will explore a set of proxy variables that were developed to relate the key factors found in these logistic regression models to crossing characteristics. This will allow the incident-specific models developed in this chapter to be adapted into a crossing-specific model that can predict the likelihood of a derailment occurring at a crossing.

CHAPTER 7: MODEL APPLICATION AND PROXY DEVELOPMENT

7.1. INTRODUCTION

Use of the joint freight-passenger train derailment likelihood model to estimate derailment probability based on crossing characteristics is described in this chapter. While Chapter 5 developed an incident-specific model, a crossing-specific model is more useful because it can be used to inform decisions about which grade crossings to upgrade to reduce derailment likelihood. To adapt the model, it is necessary to identify characteristics of the crossing that correlate with the physical properties of the incident identified in the previous chapters. These crossing characteristics are called “proxy variables” to identify their function as a bridge between the incident- and crossing-specific models. For each incident-specific variable used in the model, a crossing-specific variable (or variables) will be investigated as a proxy. Having defined these proxies, Chapter 8 will demonstrate how this information can be combined into a “derailment likelihood calculator”, used to calculate a distribution of derailment likelihoods and/or expected values of derailment likelihood.

To develop a model that can rank crossings by their derailment risk, the first step is to understand if the physical characteristics of a crossing can be used as proxy variables for the incident-specific variables used in this analysis. For example, knowing that derailment likelihood increases with increased vehicle speed is not necessarily useful unless highway vehicle speed at particular crossings can be reliably determined.

7.2. METHODOLOGY

This chapter uses data from two FRA databases: the Highway Rail Accident (HRA) database, and the U.S. DOT Grade Crossing Inventory (GCI). For each incident in the HRA database, the GCI record that was current at the time of the incident was extracted. The GCI is a historical record, and as such contains multiple records for each grade crossing. This information includes posted highway speed limit, type of crossing warning device in use, and other characteristics. Crossing-specific information such as this is subject to change over time. The GCI is supposed to be updated each time conditions change at the crossing (such as adding a gate or increased train service, etc.); therefore, when relating information from the HRA database to the GCI, it is important to select the GCI record that was in effect at the time of the HRA incident. This was done by selecting the GCI record whose effective and end dates bracket the incident date from the HRA.

To develop the proxy relationships, data were needed for a number of variables. Due to incomplete information (in the GCI, especially), the number of useable records were often less than those used to develop the incident-level models. In practice, the data in the GCI are irregularly entered and sometimes contain reporting errors. A data set for each proxy variable was developed that contained only records with complete information for the variable under study. As a result, the size of datasets differ from variable to variable.

The inventory records used may not be representative of all grade crossings. There may be reasons that inventory data were collected for some crossings and not for others. For example, more complete data may be provided by some states than others. If this is the case, there could be regional variation that is not reflected in the results. Another problem is that states might prioritize data collection at crossings with high traffic or that are otherwise of greater concern. This could bias results towards these "high impact" crossings.

7.3. ANALYSIS

7.3.1. Proxy for Highway Vehicle Speed

7.3.1.1. Deviation from Posted Highway Speed

For each record with both a valid vehicle speed and highway speed limit, a value that compares the collision speed of the vehicle to the posted highway speed limit was calculated. This value was defined as the "percent deviation from posted highway speed limit" (PDHSL), as follows:

$$\%\ deviation\ from\ posted\ highway\ speed = \frac{Vehicle\ Speed - Highway\ Speed}{Highway\ Speed}$$

The distribution of PDHSL was plotted (Figure 7.1). A value of -100% means the vehicle was stopped on the crossing. A value of 0% means the vehicle was traveling at the posted speed limit. A value of 100% means the vehicle was traveling at twice the posted speed limit. Less than 0.2% of all incidents occurred with the vehicle traveling more than twice the posted speed limit, so the chart was truncated at 100%.

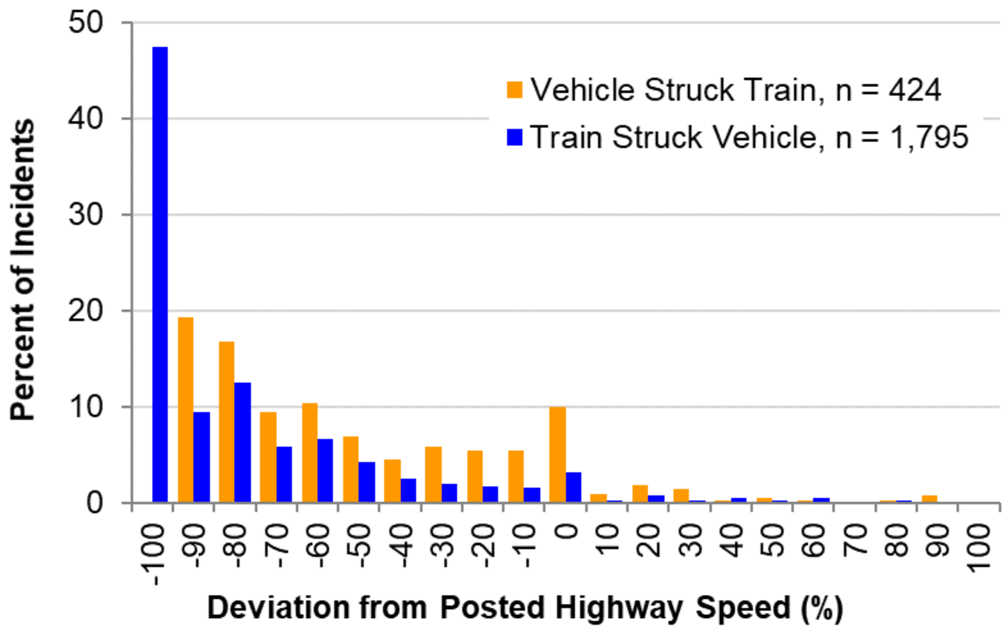


Figure 7.1: Probability mass distribution of percent deviation from posted highway speed.

In cases where the train struck the vehicle, about 47% of incidents involved a vehicle that was stopped on a crossing (Figure 7.1). Otherwise, moving vehicles were more likely to strike the train than be struck by the train, irrespective of speed.

7.3.1.2. Highway Classification and Warning Device

In addition to varying by highway speed limit and incident type, highway vehicle speed may vary by highway classification and warning device. Highway classification is based on annual average daily traffic (AADT) and is correlated with posted speed limit, so the speed distribution would be expected to vary depending on highway type. A benefit to using highway classification instead of posted speed limit is that more records in the GCI have reported values for highway classification, allowing a larger and potentially more robust dataset.

The second factor, warning device, varies from “none”, to “crossbuck only” for passive crossings, through bells, flashers, standard gates, and four-quad gates for active crossings (Appendix C). Vehicle speed might be expected to vary somewhat based on warning device type. Gates provide a physical impediment that is intended to force motorists to slow down or stop. Flashers or bells alone alert motorists to the possibility of a train but do not “force” them to stop in the same manner as gates. Motorist speeds at passive crossings with crossbucks might be even higher because the crossbuck provides no warning of an approaching train. Some motorists might not know how to respond to the sign, or become accustomed to never seeing a train at a particular crossing and ignore it. Speed distributions might follow the typical speed distributions for each highway class, which are normally distributed with a certain mean (TRB, 1998); however, for grade crossing collision speeds, none of the distributions were normally distributed (Appendix D).

The GCI database defines nine types of warning devices (WDCODE). Of these, categories 1 (“none”) and 9 (“four-quad gates”) were excluded because both of these were rare in the database, so it was not possible to develop adequate distributions for them. Furthermore, they are sufficiently different from other warning device types that it did not seem appropriate to group them together, so they were excluded from the analysis. The remaining seven categories were grouped into three larger categories based on similarities in physical characteristics and statistical distributions. The three groups are:

“Passive”: WDCODEs 2 (“other signs/signals”), 3 (“crossbucks”) and 4 (“stopsigns”)

“Other Active”: WDCODEs 5 (“special active devices”), 6 (“highway signals, wig wags, bells”) and 7 (“flashing lights”)

“Gates”: WDCODE 8 (“all other gates” (excluding four-quad)

Similarly, when considering highway class, the GCI uses 12 classes, of which six are for rural roads and six are for urban roads (Appendix E). They range from interstates to local roads for the two land use types. Again, due to problems of data scarcity and simplification, some classes were excluded from the analysis and others were grouped together. Records labeled as "interstate", both rural (HWYCLASS 1) and urban (HWYCLASS 11) were excluded. There were little data in those categories since grade crossings on true interstate highways are rare. The remaining groups are rural arterial, rural collector, rural local, urban arterial, urban collector, and urban local.

This yields a 3 x 7 matrix of speed distributions for each of two incident types (VST and TSV-M), for a total of 42 distributions. The challenge was to balance the risk of overfitting the data versus accurately representing the probable incident speed distribution. With these categories, each cell in the speed distribution has a minimum of 100 records.

7.3.2. Proxy for Train Speed

7.3.2.1. Deviation from Timetable Speed

Similar to the analysis of vehicle speed, for each record with both a valid train speed and timetable speed, a value was calculated that compares the collision speed of the train to the timetable speed on the rail line. This value, which will be called "percent deviation from timetable speed" (PDTTS) is defined as:

$$\%\ deviation\ from\ timetable\ speed = \frac{Train\ Speed - Timetable\ Speed}{Timetable\ Speed}$$

Figure 7.2 shows the distribution of PDTTS. A value of -100% means the train was stopped on the crossing. A value of 0% means the train was traveling at the timetable speed. A value of 100% means the train was traveling at twice the timetable speed. While some trains (less than 1%) were reported to be traveling more than twice the timetable speed, these records were excluded from analysis because they are probably the result of reporting errors. For the most part, these records reported a timetable speed of 0 or 1 mph, which is implausible.

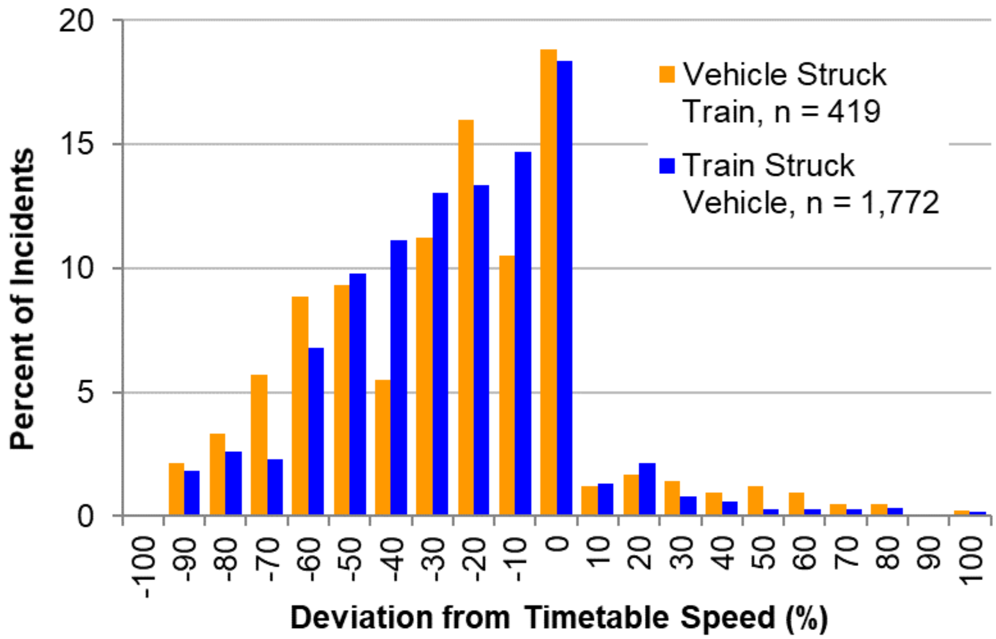


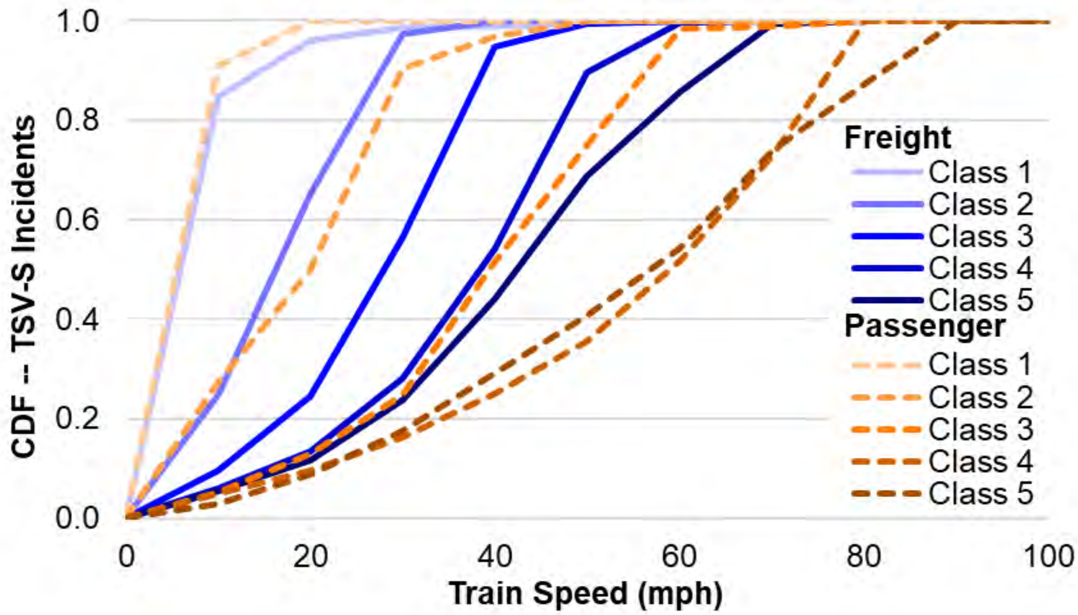
Figure 7.2: Probability mass distribution of percent deviation from timetable speed.

Most trains were traveling near the timetable speed when the collision occurred (Figure 7.2). Interestingly, no trains were stopped on the crossing at the time of collision. About 8% of incidents involved a train traveling in excess of the timetable speed. If true, this represents about 150 over-speed incidents over 20 years of operation. These incidents are more likely to be occurring on the lower FRA track classes, where twice the timetable speed is still not very high.

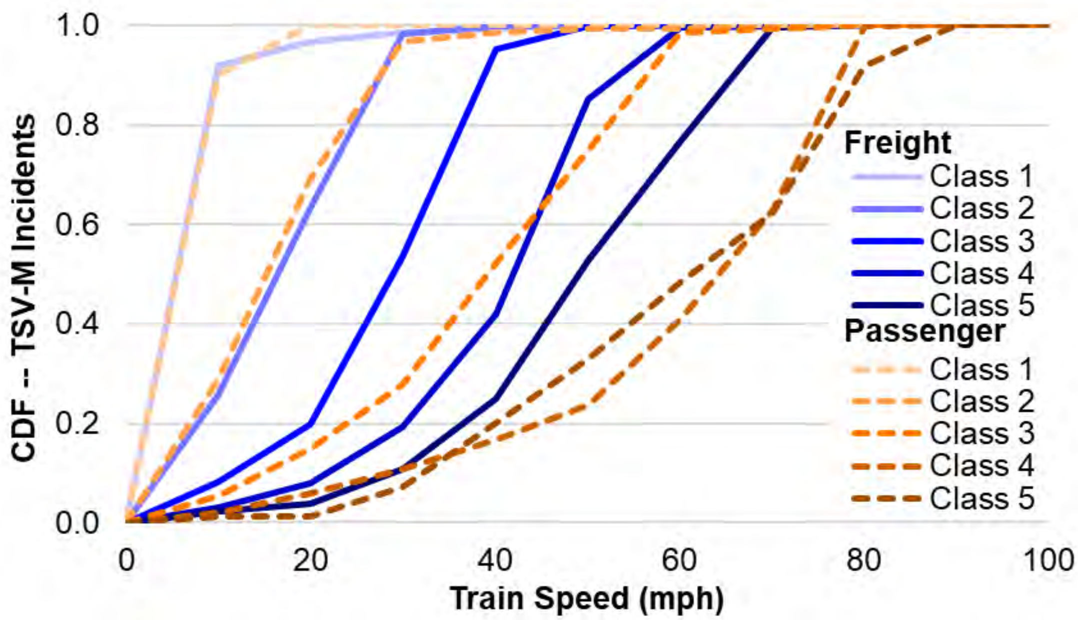
While some of these events are confirmed over-speed incidents, many are probably reporting errors. This could be determined on a case-by-case basis if data regarding the particular timetable in effect at the time of the incident were obtained from the railroad.

7.3.2.2. As a Function of Track Class and Equipment Type

To further refine the predictive ability of the model, train speed was examined as a function of FRA incident type, track class and train type (freight vs. passenger). Train collision speed data from the HRA database were sorted into ten-mile-per-hour bins and plotted to determine the distribution of train speeds (Figure 7.3, Table G.1) (see Table 1.1 for track class speeds). Attempts to fit common statistical distributions to these data were inconclusive; therefore, an empirical distribution was used based on the 20-year study data.



(a)



(b)

Figure 7.3: CDF of train speeds, showing freight and passenger train incidents for FRA track classes 1 through 6 for (1) TSV-S incidents and (b) TSV-M incidents. Note: The speed distribution for class 1 passenger trains was developed using reported maximum timetable speed instead of track class.

7.3.3. Proxy for Highway Vehicle Type

7.3.3.1. Percent Truck Traffic

It was hypothesized that crossings with greater truck traffic are more likely to have train-truck collisions. This was tested using the GCI data for the percentage truck traffic at crossings to quantify the number of trucks expected.

Figure 7.4 shows the distribution of percent truck traffic (PTT) for cases that did or did not involve a large highway vehicle. A t-test comparing the two distributions showed a small, but statistically significant ($p = 0.0113$, $\alpha = 0.025$) difference between the likelihood of large vehicle involvement based on percent truck traffic at the crossing. Specifically, the likelihood that an incident will involve a truck is greater for crossings with a higher percentage of truck traffic.

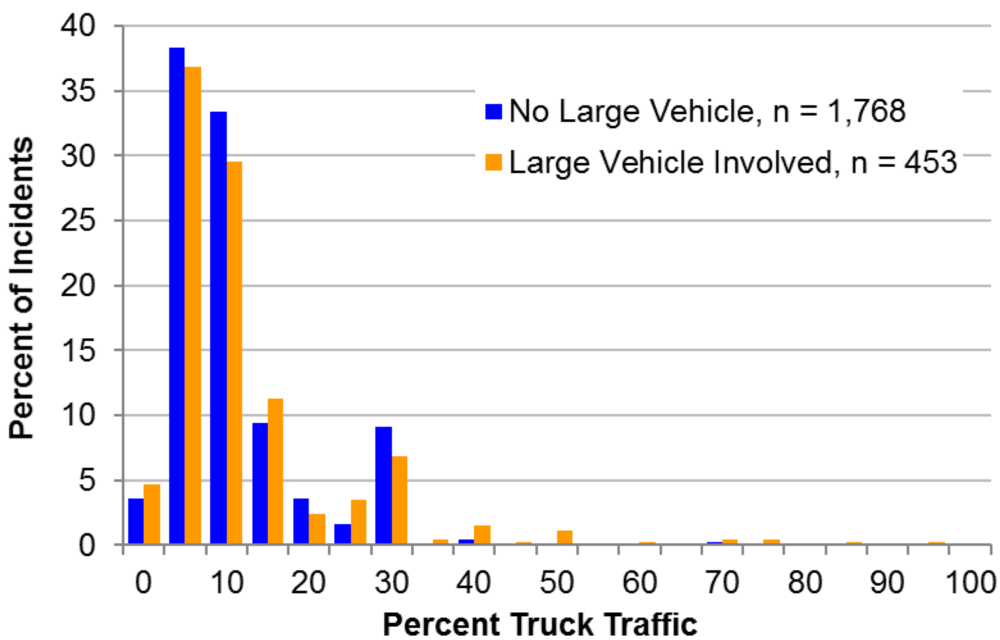


Figure 7.4: Distribution of Percent Truck Traffic for incidents involving cars and incidents involving trucks.

Overall, it appears that there is a relationship between percent truck traffic and the likelihood of an incident involving a truck; however, the relationship is not strong enough to help predict this likelihood. Attempts were made to develop a series of models of vehicle size using logistic regression (since vehicle size has been defined here as a binary variable, large or small). This model would predict the likelihood of an incident at a specific crossing involving a large highway vehicle. As input to this model, highway class, warning device type, train type (freight/passenger) and incident type were considered, in addition to percent truck traffic and annual average daily traffic (AADT). For this analysis, the same groups for highway type and warning device type were used as in the highway speed analysis presented in Section 7.3.1 (Appendices C and E).

No robust logistic model was identified based on any combination of the factors listed in the previous paragraphs. They produced results that were only slightly better than random guessing, with AUC values between 0.50 and 0.55, and large Hosmer-Lemeshow test values (see Chapter 3 for definitions of these values). Multiple models based on the first four variables were also tried, but did not provide better results. This was followed by location testing that compared PTT distributions for large and small vehicles, divided by highway class (HC), warning device type (WD), train type (TT), and incident type (IT). This produced 108 "cells" (6 HC x 3 WD x 2 TT x 2 IT). For 77 of these cells, no statistically significant difference was detected (alpha = 0.10.) Three cells showed that incidents involving small vehicles occurred at crossings with higher PTT; 22 cells showed that incidents involving large vehicles occurred at crossings with higher PTT. The remaining six cells had insufficient data to draw a conclusion.

Only 5.3% of incidents occurred on roads with more than 20% truck traffic, and only 1.6% of incidents occurred on roads with greater than 30% truck traffic. Some of the modeling problems could stem from the lack of data at the high end of the PTT range, and better models might result from truncating the dataset. All data points with more than 20% PTT were removed, and the modeling approaches repeated as described above. The results were similarly ineffective at predicting the size of the vehicle involved in the collision.

Therefore, overall it appears that the effect of PTT on the likelihood of an incident involving a large vehicle is indeterminate. More importantly, even for cases where PTT had an effect (according to the tests of location), a logistical model developed using the factors discussed above did not successfully identify large-vehicle-involved incidents.

7.3.3.2. Analysis Based on Other Crossing Characteristics

A simple but robust approach for modeling large-to-small vehicle ratio divided the data by highway class, warning device type, train type (freight/passenger) and incident type (VST/TSV-M/TSV-S). Ideally, knowledge of the ratio of truck traffic to car traffic should be used to predict the type of highway vehicle involved in an incident; however, due to the limitations of available data this was not feasible. The ratios of incidents involving large and small vehicles were plotted according to the four factors (Figure 7.5, Table G.2). A sample of the figures are shown here; the figures for all categories can be found in Appendix F.

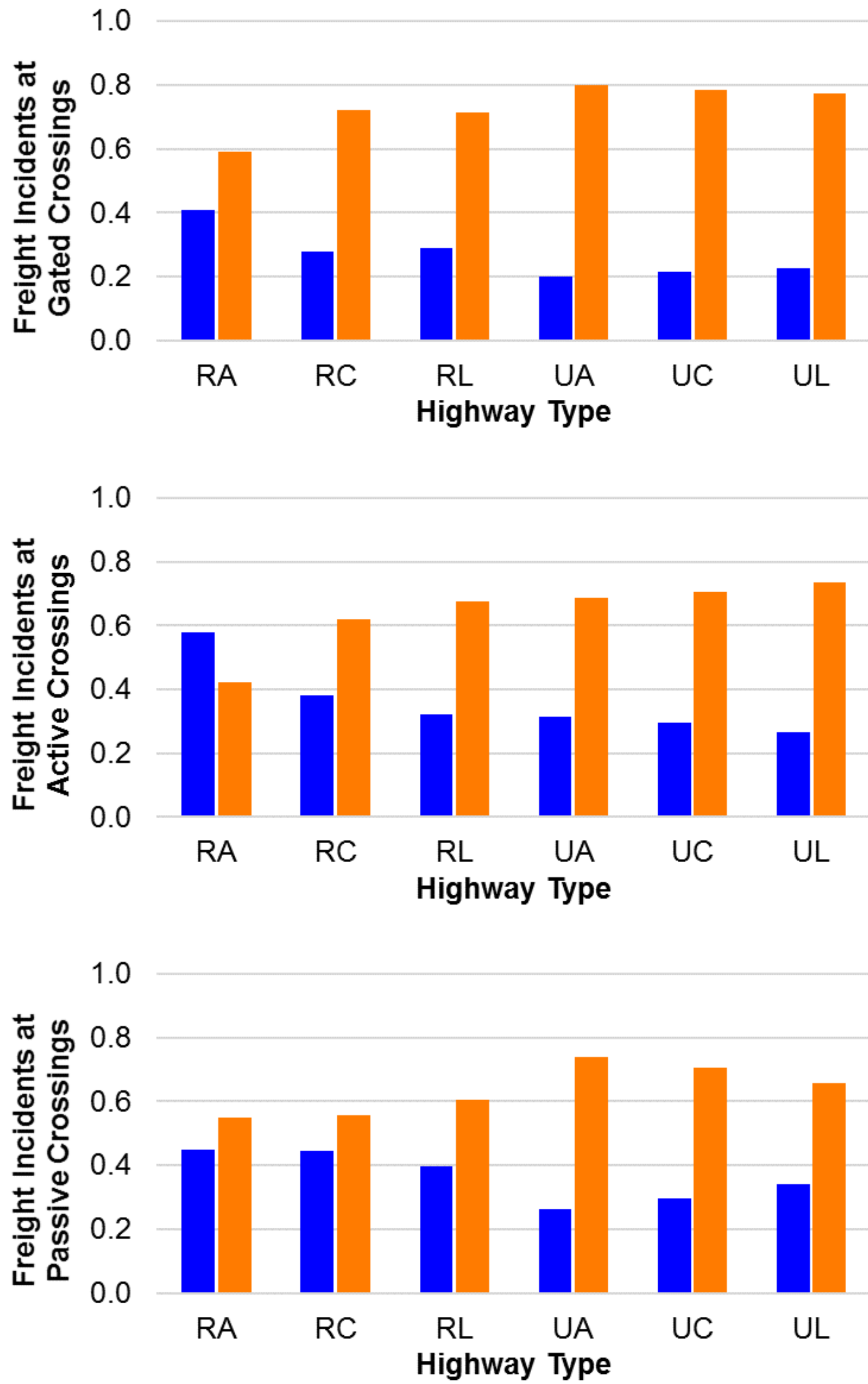


Figure 7.5: Ratio of large to small vehicles in TSV-M incidents at (top) gated crossings, (middle) active crossings and (bottom) passive crossings for freight trains.

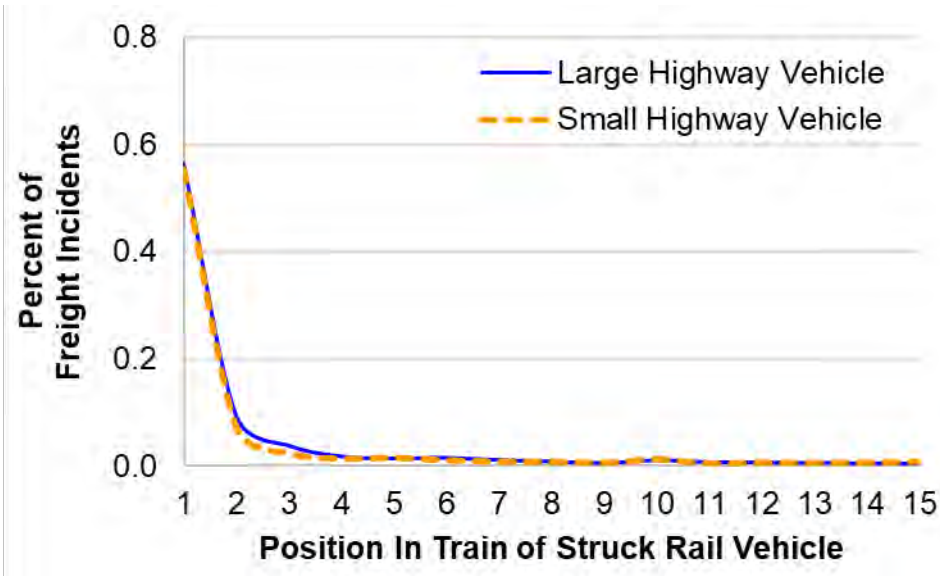
7.3.4. Proxy for Equipment Class

Equipment class is a complicated proxy to develop as it relies on several factors. Ultimately, three factors were used: the likelihood of an incident being a freight or passenger train, the distribution of position-struck-in-train, and the distribution of the number of locomotives. As equipment class is only a factor in derailment likelihood for VST incidents, proxies were developed based only on data for these incidents.

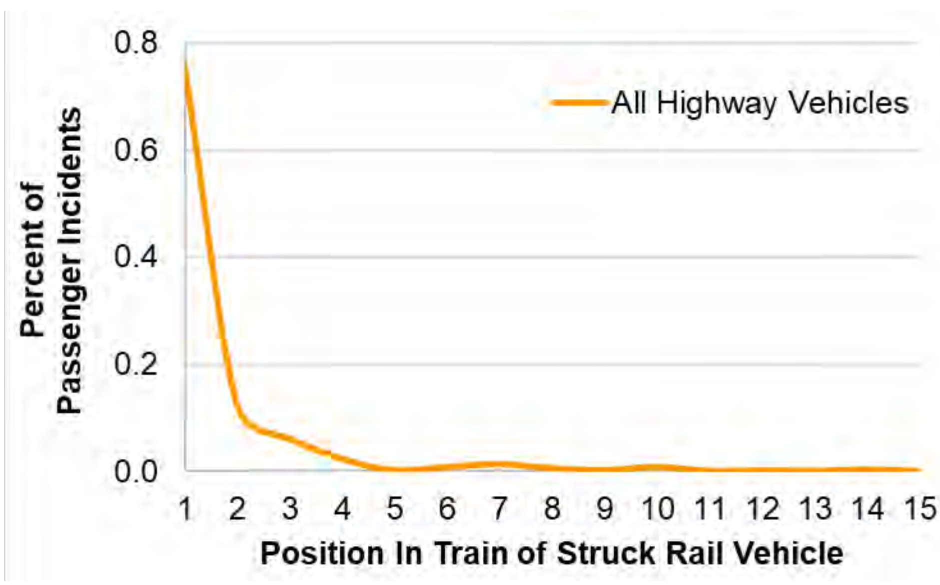
The GCI database has two variables useful to this analysis: PASSCNT and TOTALTRN. PASSCNT is the number of passenger trains that use the crossing each day. TOTALTRN is the total number of trains that use the crossing each day. The percentage of trains using a crossing that are passenger trains is PASSCNT/TOTALTRN.

In order to know if an incident is likely to involve a locomotive or a railcar two sets of distributions were developed, a distribution of where in the train a highway vehicle is likely to hit in a VST incident ("distribution 1") (Figure 7.6, Table G.3), and a distribution of the number of locomotives in a train consist ("distribution 2") (Figure 7.7, Table G.4).

For distribution 1, the data were first separated by train type (freight or passenger). For freight train VST incidents, over 50% of highway users struck the first rail vehicle in the consist (Figure 7.6a). For passenger train VST incidents, over 70% of highway users struck the first rail vehicle (Figure 7.6b). This difference may be due to the fact that passenger trains occupy a crossing for a shorter period of time than freight trains, therefore the freight train's exposure to VST incidents not involving the first rail vehicle is greater.

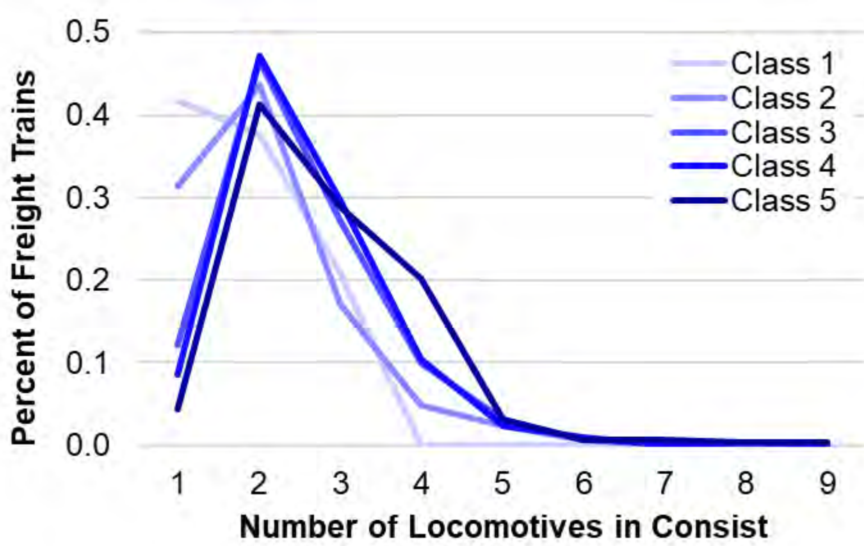


(a)

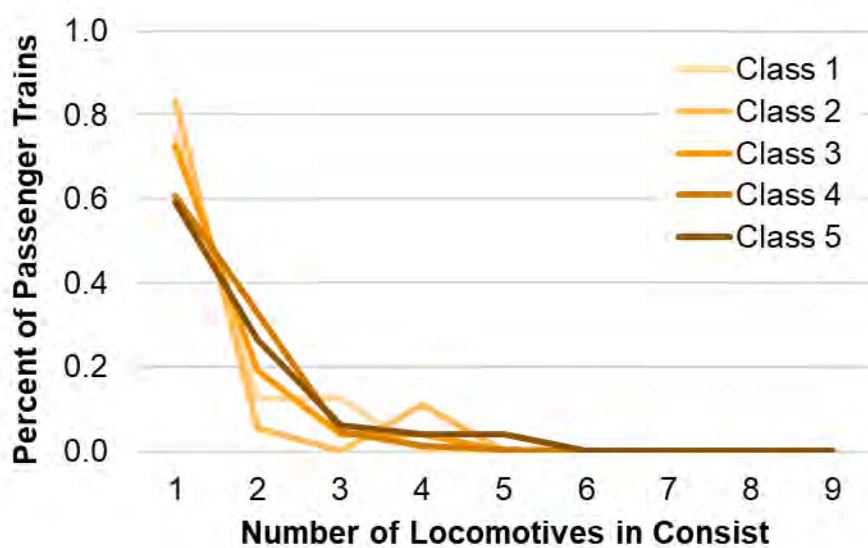


(b)

Figure 7.6: Position in train of struck rail vehicle for VST incidents involving (a) freight and (b) passenger trains.



(a)



(b)

Figure 7.7: Distribution of number of lead locomotives on train consist for (a) freight and (b) passenger trains.

The data were further separated by highway vehicle type (large or small). Due to differences in vehicle characteristics and driver behavior, it seemed possible that the distribution of position-struck-in-train might differ between large and small vehicles. However, since it was unclear which vehicle type would be more likely to collide with a train further back in the consist, a two-tailed test was used to compare the distributions.

For freight train incidents, there was a small, but statistically significant difference between the position-struck-in-train distributions for large and small vehicles (Figure 7.6a) ($Pr > KSa < 0.0001$), with small vehicles on average striking further back in the train than large vehicles. For passenger train incidents, there was no statistically significant difference ($Pr > KSa = 1.000$) between the distributions for small and large vehicles (Figure 7.6b). Therefore, two separate distributions for the freight model were used, and one distribution was used for the passenger model. For both types of train, there was a rapid initial decrease to the function, but by the tenth rail vehicle in the consist the likelihood of any individual vehicle being struck is essentially random with a likelihood close to zero.

For distribution 2, since there is no reason to expect that the number of locomotives in a consist would affect incident type, all grade crossing incidents were used to develop the distributions. Two sets of distributions were developed, one for freight (Figure 7.7a) and one for passenger (Figure 7.7b). It was assumed that all locomotives in the consist were placed at the front of the train. This is a simplified assumption, since some locomotives may be placed elsewhere in the train. The HRA database provides information about the number of locomotives in the train, but does not provide any information about their position. The REA database provides slightly more information, with fields for locomotives at the front, middle and rear of the train; however, it is not possible to know exactly where the middle locomotives are placed. Furthermore, due to the necessity of also accounting for train length distributions, trying to model an accurate distribution of locomotives required dividing the data into very small groups, which made it difficult to draw statistically significant conclusions.

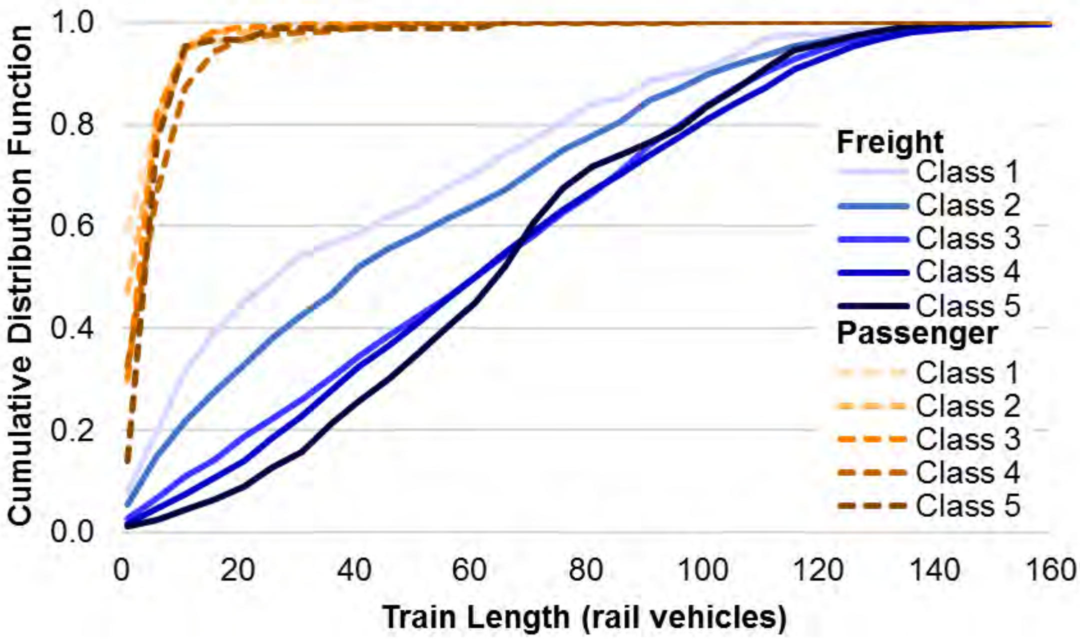
Since track class correlates with speed, and speed of a train correlates roughly with the amount of power required to move the train, the number of locomotives could vary with track class. Therefore, separate distributions were developed for each track class. Trains traveling on higher-class track generally had more locomotives (Figure 7.7).

It should be noted that it is uncommon for passenger trains to have large numbers of locomotives. The REA database identified seven passenger trains with six or more locomotives. On inspection, these appeared to be electric multiple unit (EMU) or diesel multiple unit (DMU) trains, operated by commuter rail systems. According to the FRA, EMUs and DMUs should be counted as railcars for the purposes of accident reporting; however, due to user error these are sometimes identified as locomotives instead. Since EMUs and DMUs are more similar in weight and structure to a railcar, any record that incorrectly identified them as locomotives was removed from the dataset. The REA database identified an additional 12 passenger trains with four or five locomotives. These consisted of long-distance Amtrak trains, mostly in parts of the U.S. that likely have steep grades and may require the additional power. Combining distributions 1 and 2 enables prediction of how likely it is that the rail vehicle struck in a VST incident is a locomotive, based on track class. The distributions for freight and passenger are combined based on the ratio of freight and passenger traffic at the grade crossing.

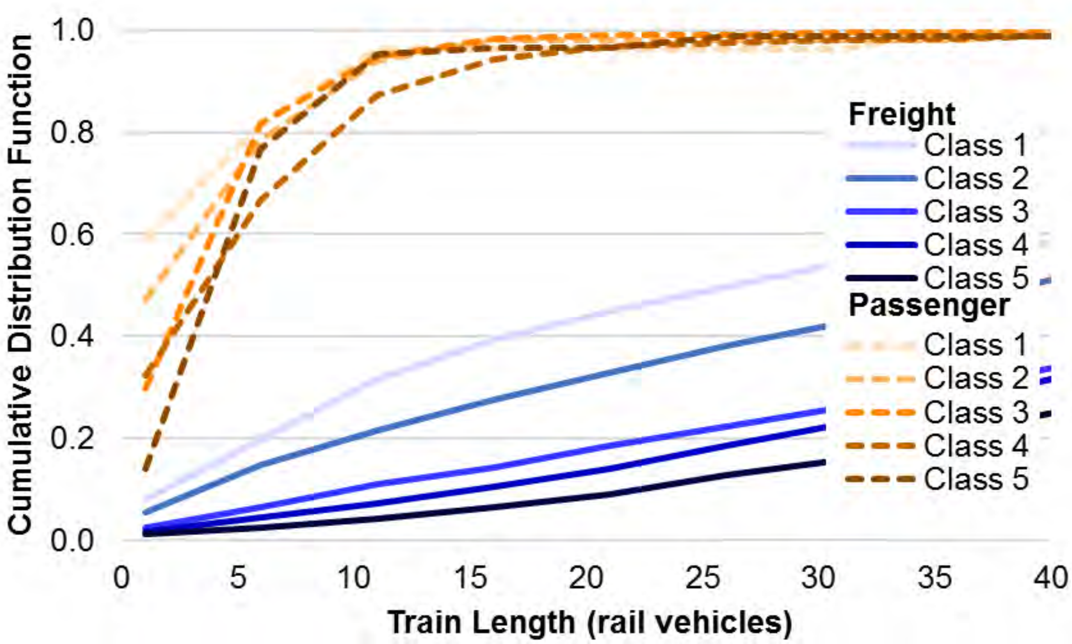
7.3.5. Proxy for Train Length

Train length is an input for the TSV-S model. Train length differs between freight and passenger trains as well as between track classes. In the U.S., freight trains can be in excess of 100 cars long, and with limited exceptions passenger trains rarely exceed 20 cars. Freight train length could be expected to vary with track class because railroads maintain track to higher standards if they have high enough traffic volumes to justify the investment in higher speed operation.

For the track classes in the database, the distribution of train lengths was determined for freight and passenger trains (Figure 7.8, Table G.5). Only track classes 1 through 5 were used; class 6 track and excepted (class “X”) track were excluded due to sparse data. As with the highway speed analysis, these data could not be fitted to common statistical distributions; therefore an empirical distribution based on the 20 years of data in the study was used.



(a)



(b)

Figure 7.8: Distribution of train lengths by track class and consist type for (a) all trains and (b) trains shorter than 40 rail vehicles.

7.3.6. Proxy for Incident Type

There may be a number of factors that influence whether a crossing is more likely to experience a VST or TSV incident, including human-factors aspects that are outside the scope of this work. A simple proxy based on warning device type in use at the crossing was used (Table 7.1). The data indicated that the incident type ratios vary depending on the type of warning device at a crossing. For example, VST incidents are substantially less frequent at gated crossings compared to active or passive crossings because gates provide a more obvious physical barrier. Subsequent work could refine this element through a better understanding of which crossing characteristics affect incident type.

Table 7.1: Incident Type by Warning Device Type

	Warning Device Type		
Incident Type	Gates	Other Active	Passive
VST	18.8%	32.1%	26.1%
TSV-S	51.1%	24.4%	23.5%
TSV-M	30.0%	43.5%	50.4%

7.4. DISCUSSION

Ideally, one would use crossing-specific data for the proxy variables in lieu of the proxy variable distributions developed in this chapter. This would reduce the need for assumptions and in theory provide a more accurate estimate of derailment likelihood at a crossing. It would also limit the risk that model users might use the proxy variables inappropriately or based on out-of-date information.

However, in most cases it is not feasible (or even possible) for a user to obtain the necessary data based on crossing observation Any of the factors that vary based on incident type (highway vehicle speed, train speed, highway vehicle type) cannot currently be provided by users, because even if they can provide the distribution of the factor at the crossing, it is not clear how observed distributions relate to incident-type-specific distributions. If future researchers can determine these relationships based on field studies or other techniques, then the calculator can be adapted to incorporate user-provided distributions.

A good example of this is highway vehicle type. It seems logical that the best predictor of LGVEH would be percent truck traffic (PTT). However, PTT does not appear to be an accurate predictor of LGVEH, as quantified by my attempts to conduct multivariate analysis and generate regression models, as explained in Section 7.3.3.1. This could be a problem with the data in that PTT in the GCI might not represent the actual ratio of large to small vehicles at a crossing. It is my understanding that these counts are typically generated by an observer who monitors the road for a period of 24 hours or less. This means that fluctuations in PTT over time (both within a week and over the course of a year) are not accounted for. There is no guarantee that the observation period is representative of other days/times. This could account for the difficulty in developing statistical relationships.

Even if the factor does not vary with incident type, there are other considerations. Accounting for equipment class in an incident requires a distribution of position-struck-in-train (Figure 7.6) that is necessarily based on historical data and not observation.

Train length is the only factor for which users could provide custom data, as it is not dependent on incident type or historical observation.

7.5. CONCLUSIONS

The distributions developed in this chapter can be combined with the statistical regression models developed in Chapter 6 to predict a probability distribution for a grade crossing incident-caused derailment to occur at a specific grade crossing based on its characteristics. This probability distribution can also be simplified into a point estimate, which simplifies comparison but ignores consideration of variability. This methodology will be discussed in Chapter 8.

CHAPTER 8: DEVELOPMENT OF AN EXCEL TOOL TO EVALUATE DERAILMENT LIKELIHOOD AT HIGHWAY-RAIL GRADE CROSSINGS

8.1. INTRODUCTION

In this chapter, the "derailment likelihood calculator" is developed. It consists of an Excel spreadsheet combining the proxy variables discussed in Chapter 7 with the joint freight-passenger derailment model from Chapter 6. The spreadsheet can be used to calculate a distribution of derailment likelihoods, as well as an expected value of derailment likelihood and other estimates.

8.2. MATHEMATICAL BACKGROUND

The distributions from Chapters 6 and 7 can be combined using probability concepts. A joint distribution can be developed combining all variables and their distributions.

For a function $g(x)$ of continuous random variable x, with probability density function $f(x)$, the expected value or expectation of that function is defined as (Modarres et al., 2009)

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx \tag{1}$$

Analogously, for discrete distributions $f(x_i)$ with probability densities $pr(x_i)$, the expected value is defined as

$$E[g(x)] = \sum\nolimits_{i=1}^{k} g(x_i)pr(x_i) \tag{2}$$

This can be generalized to joint probability distributions with multiple variables x_n,

$$E[h(x_1, x_2, \dots, x_n)] = \sum_{x_1}\sum_{x_2} \dots \sum_{x_n} h(x_1, x_2, \dots, x_n)pr(x_1, x_2, \dots, x_n) \tag{3}$$

Furthermore, algebraically, for two functions f(x) and g(y),

$$E[f(x) + g(y)] = E[f(x)] + E[g(y)] \tag{4}$$

8.3. CALCULATOR FUNCTIONALITY

The calculator consists of a Microsoft Excel workbook with 14 spreadsheets. The first sheet, “Calculator”, is edited by users to run the calculation for an individual crossing. The second sheet, “Data”, gives users the ability to process a batch of crossings. The third sheet, “Logistic Models”, shows the three logistic models that were developed to predict derailment likelihood (one each for VST, TSV-S and TSV-M). This is a reference for users to understand the underlying equations.

Four sheets perform the calculations that produce the estimates of derailment probability, one each for VST, TSV-S and TSV-M, plus a final “Combination” sheet that combines the three distributions based on the expected ratio of VST:TSV-S:TSV-M incidents. The “Combination” sheet performs the calculations for the results that are shown on the “Calculator” sheet.

The remaining seven sheets contain the distributions developed for each of the proxy variables discussed in Chapter 7. These are used for calculations on the VST, TSV-S and TSV-M spreadsheets.

Table 8.1 shows the relationships between the base variables of the regression models, the proxy variables, and required inputs. Each of the calculation spreadsheets will be discussed.

Table 8.1: Relationship between base variables, proxies and calculator inputs

Base Variable	Proxy	Inputs	Input Type	Models
Incident Type (TYPACC)	Historical incident type ratio	Warning device type (WDCODE) groups*	User selected	Combination model
Vehicle Speed (VS)	Vehicle speed distributions	WDCODE groups*	User selected	VST, TSV-M
		Highway classification (HWYCLASS) groups**	User selected	
Large Vehicle (LV)	Large-to-small-vehicle ratios (LGSM)	TYPACC	Proxy	VST, TSV-S, TSV-M
		FRTPAX	Proxy	
		WDCODE groups*	User selected	
		HWYCLASS groups**	User selected	
Equipment Class (EC)	Freight-to-passenger-train ratio (FRTPAX)	Daily number of passenger trains (PASSCNT)	User input	VST
		Total daily number of trains (TOTALTRN)	User input	
	Positon-in-train of struck rail vehicle distributions	LGSM	Proxy	
	Number of locomotives in the train distributions	Track class (TRKCLAS)	User selected	
Train Speed (TS)	Train speed distributions	FRTPAX	Proxy	TSV-S, TSV-M
		TRKCLAS	User selected	
Train Length (TL)	Train length distributions	FRTPAX	Proxy	TSV-S
		TRKCLAS	User selected	

*As defined in Appendix C
** As defined in Appendix E

8.3.1. Calculator Input/Output Spreadsheet

The calculator requires the user to input five values to calculate the conditional probability of derailment, $p(D|I)$ (Figure 8.1). All values can be found in the FRA's Grade Crossing Inventory (GCI), if the records for the crossing there are complete and up-to-date (a major caveat, because data quality and completeness in the GCI vary considerably from state to state).

Conditional Probability of Derailment [P(D|I)] Calculator		
Enter Crossing Factors		
Grade Crossing Warning Device Type	Gates	
Highway Class	Urban Arterial	
Passenger Trains per Day (PASSCNT)	6	(numeric)
Total Trains per Day (TOTALTRN)	25	(numeric)
FRA Track Class (RRCLAS)	5	
Results		
P(D|I) -- Expected Value	0.0828	
P(D|I) -- 5th percentile	0.0783	
P(D|I) -- 25th percentile	0.0870	
P(D|I) -- 50th percentile	0.1565	
P(D|I) -- 75th percentile	0.2519	
P(D|I) -- 95th percentile	0.3437	

Figure 8.1: Input/Output view of Excel spreadsheet "Calculator"

Three of the five factors are limited to a predefined set of values so a drop-down menu is used. These factors are grade crossing warning device type, highway classification, and FRA track class. Users can select from three categories of warning device type: gates, other active, or passive. They can select from six categories of highway classification. These differ from the classifications given in the GCI; the categories are simplified, as shown in Appendix E. Lastly, users select an FRA track class between 1 and 5. The calculator does not include default values for excepted track (class "X") or for track classes above 5.

The total number of trains per day and the number of passenger trains per day must also be provided by the user. These values are used to determine percent passenger trains at the crossing. The total number of trains per day must be greater than zero and must be greater than the number of passenger trains per day. The spreadsheet verifies that the input is a valid number.

The calculator provides two types of output. First, it provides six estimates of the conditional probability: the expected value of $p(D|I)$, using the expectation of the functions (explained mathematically in subsequent sections), as well as the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentile of the probability function. These point estimates describe the overall shape of the conditional probability function, The 95^{th} percentile in particular provides an upper bound to the $p(D|I)$ function but truncates the long tail.

The methodology developed in this dissertation relies on distributions for each proxy variable. Therefore, to provide users with a more comprehensive understanding of the probability of a grade crossing experiencing a derailment, a distribution of all possible $p(D|I)$ values for the given crossing characteristics (Figure 8.2) is also provided. The mathematics behind the development of this distribution is discussed in more detail in a subsequent section.

The distribution is important because it shows that, while the highest value of $p(D|I)$ predicted by the calculator may be much greater than the expected value or 95^{th} percentile, it is also much less likely to occur. The goal of combining the point estimates with the distribution figure is to provide users with perspective about the probability of derailment at a crossing. It is important to understand that there is variability in the estimate, without overstating the likelihood of an incident occurring.

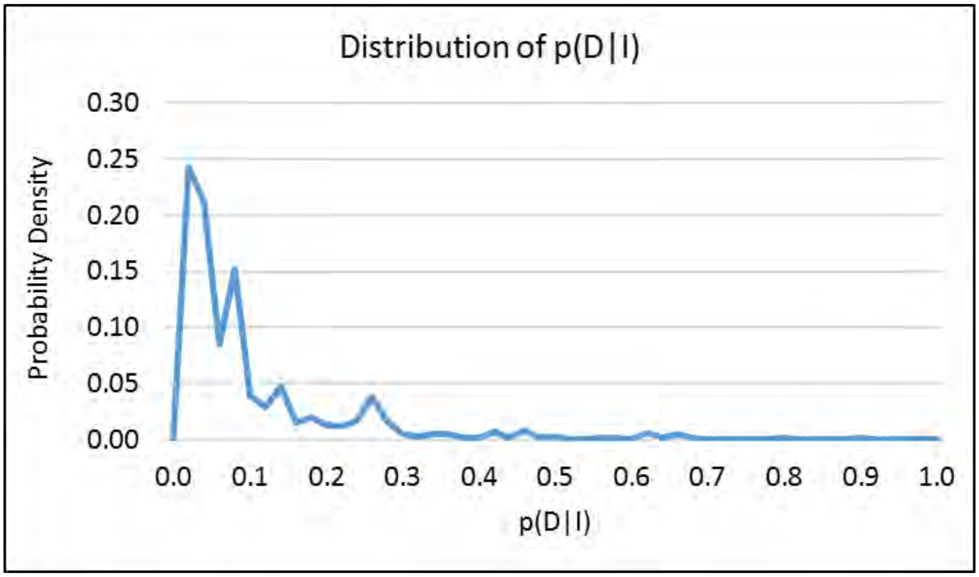


Figure 8.2: Sample output figure of the distribution of conditional probabilities.

8.3.2. Incident Type Calculations

As components of the overall derailment calculator, each of the three incident-type-specific equations were first evaluated with the corresponding model variable distributions. This process is described below.

8.3.2.1. VST Calculation

The VST calculation relies on three model variables: vehicle speed, highway vehicle size, and equipment class. These are represented by five proxy variables: vehicle speed distributions, the expected ratio of large to small vehicles, the ratio of freight to passenger trains, the position-in-train of the rail equipment, and the number of locomotives in the consist.

The appropriate vehicle speed distribution is selected based on the warning device type and highway classification. The large-to-small-vehicle ratio is also selected based on these two inputs, as well as the freight-to-passenger-train ratio. The final three proxies are selected based on the freight-to-passenger-train ratio, the large-to-small-vehicle ratio, and FRA track class.

Two approaches were used to combine this information and develop an estimate of the $p(D|I)$ value for a crossing. The first is to calculate the expected value of the x_{VST} equation and then use this value to calculate the expected p_{VST}. The second is to propagate the distributions of the proxy variables through the x_{VST} and p_{VST} equations to preserve the distribution of probability values.

The VST spreadsheet first determines the likelihood of each rail vehicle in the train being struck. Different distributions are used for freight and passenger trains; these distributions were developed in Chapter 7. For example, about 56% of VST freight train incidents involve the first rail vehicle in the train, 9% involve the second vehicle, 4% involve the third, etc. (Figure 7.6). Next, the number of locomotives in the train is accounted for, which varies with track class. For example, approximately 12% of freight trains on class 3 track have one locomotive, 46% have two locomotives, 27% have three locomotives, etc. (Figure 7.7a).

By combining the two distributions, the likelihood of a collision involving freight locomotives, freight cars, passenger locomotives, or passenger cars (the factor "EC") is calculated. These likelihoods are then multiplied by the probability distribution of vehicle speeds ("VS"), and the probability of highway vehicle size ("LV"). This produces a matrix of weights indicating how likely an incident with that combination of factors is to occur. This matrix is referred to as $pr_{VST}(VS, LV, EC)$.

Next, the logistic regression equation for VST incidents was evaluated for the possible range of vehicle speeds (as determined by highway classification and warning device type), and for each equipment class and vehicle size. This produced another matrix, with probability values for each combination of factors. This matrix is referred to as $f_{VST}(VS, LV, EC)$.

Multiplying these two matrices together and then summing the values produces the expectation of the regression equation:

$$E[f_{VST}(VS, LV, EC)] = \sum_{VS} \sum_{LV} \sum_{EC} f_{VST}(VS, LV, EC) pr_{VST}(VS, LV, EC)$$

This produces the expected value of the conditional probability of derailment for VST incidents. It is also used to develop the distribution of potential $p(D|I)$ values (Figure 8.3). For each value of $f_{VST}(VS, LV, EC)$, all corresponding values of $pr_{VST}(VS, LV, EC)$ are summed (since multiple combinations of factors could produce the same value of $f_{VST}(VS, LV, EC)$). This produces a probability distribution function illustrating the likelihood of each value of the conditional probability of derailment.

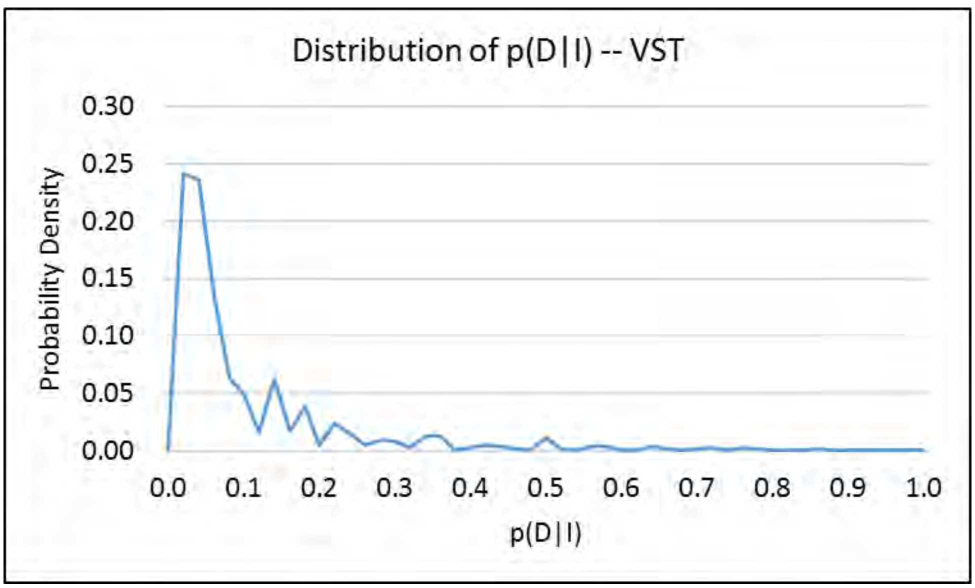


Figure 8.3: Example distribution of p(D|I) values for VST incidents.

8.3.2.2. TSV-S Calculation

A similar calculation process is undertaken for TSV-S incidents. The TSV-S calculation relies on three model variables: train speed, train length, and highway vehicle size. These are represented by a total of three proxy variables: train speed distributions, train length distributions, and the expected ratio of large to small vehicles.

The appropriate train speed distribution and train length distribution are selected based on the freight-to-passenger-train ratio and track class. The large-to-small-vehicle ratio is selected based on the freight-to-passenger-train ratio, warning device type and highway classification.

The expected value of the x_{TSV-S} equation is calculated, then this value is used to calculate the expected p_{TSV-S} as well as propagating the distributions of the proxy variables through the x_{TSV-S} equation to preserve the distribution of probability values.

The TSV-S spreadsheet multiplies the train speed distribution (the factor "TS") by the appropriate train length distribution ("TL"), and adds in the effect of highway vehicle size ("LV"). This produces a matrix of weights indicating how likely an incident with that combination of factors is to occur. This matrix is referred to as $pr_{TSV-S}(TS, TL, LV)$.

Next, the logistic regression equation for TSV-S incidents is evaluated for the possible range of train speeds, train lengths and vehicle size. This produces another matrix, with probability values for each combination of factors. This matrix is referred to as $f_{TSV-S}(TS, TL, LV)$. Multiplying these two matrices together and then summing the values produces the expectation of the regression equation:

$$E[f_{TSV-S}(TS, TL, LV)] = \sum_{TS} \sum_{TL} \sum_{LV} f_{TSV-S}(TS, TL, LV) pr_{TSV-S}(TS, TL, LV)$$

As for VST incidents, this formula is used to produce the expected value of the conditional probability of derailment for TSV-S incidents, as well as the probability distribution of potential $p(D|I)$ values (Figure 8.4).

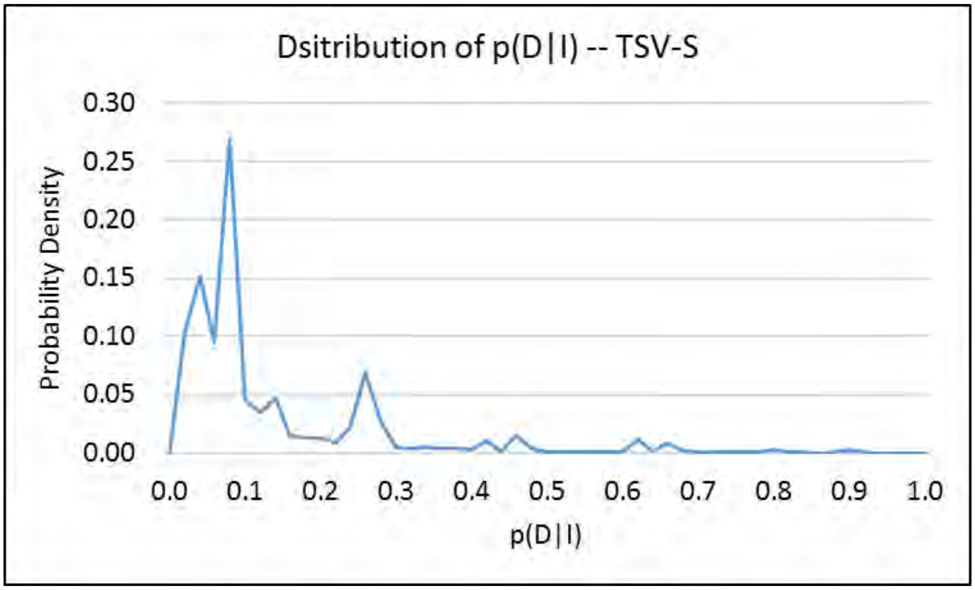


Figure 8.4: Example distribution of p(D|I) for TSV-S incidents.

8.3.2.3. TSV-M Calculation

The TSV-M calculation relies on three model variables: vehicle speed, train speed, and highway vehicle size. These are represented by a total of three proxy variables: vehicle speed distributions, train speed distributions, and the expected ratio of large to small vehicles.

The appropriate vehicle speed distribution is selected based on the warning device type and highway classification, while the train speed distribution is selected based on the freight-to-passenger-train ratio and track class. The large-to-small-vehicle ratio is also selected based on the freight-to-passenger-train ratio as well as the warning device type and highway classification.

The expected value of the x_{TSV-M} equation is calculated, then used to calculate the expected p_{TSV-M} as well as propagating the distributions of the proxy variables through the x_{TSV-M} equation to preserve the distribution of probability values.

The TSV-M spreadsheet multiplies the vehicle speed distribution (the factor "VS") by the train speed distribution ("TS"), and adds in the effect of highway vehicle size ("LV"). This produces a matrix of weights indicating how likely an incident with that combination of factors is to occur. This matrix is referred to as $pr_{TSV-M}(VS, TS, LV)$.

Next, the logistic regression equation is evaluated for TSV-M incidents for the possible range of train speeds, train lengths and vehicle size. This produces another matrix, with probability values for each combination of factors referred to as $f_{TSV-M}(VS, TS, LV)$.

Multiplying these two matrices together and then summing the values produces the expectation of the regression equation.

$$E[f_{TSV-M}(VS, TS, LV)] = \sum_{VS} \sum_{TS} \sum_{LV} f_{TSV-M}(VS, TS, LV) pr_{TSV-M}(VS, TS, LV)$$

This produces the expected value of the conditional probability of derailment for TSV-M incidents, as well as the probability distribution of potential $p(D|I)$ values (Figure 8.5).

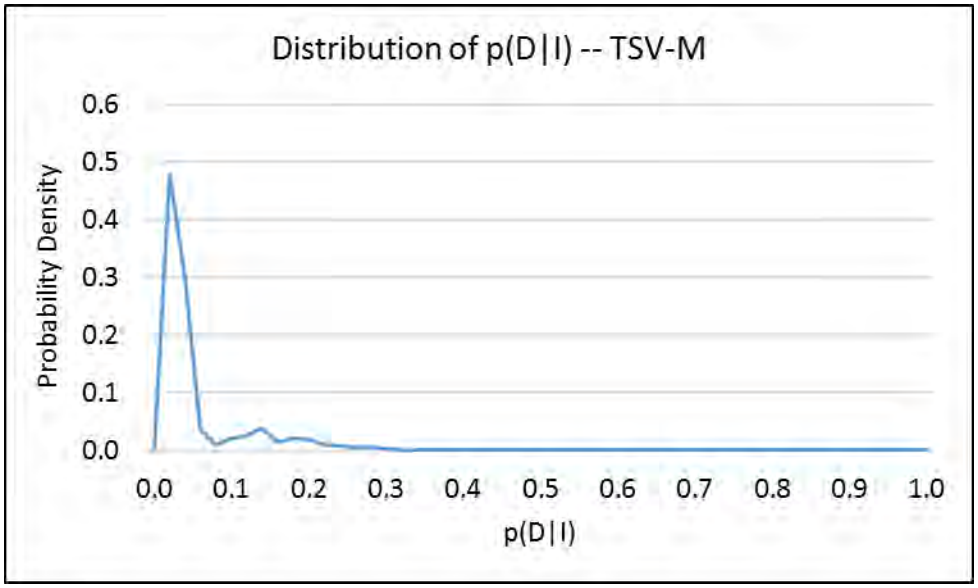


Figure 8.5: Example distribution of $p(D|I)$ for TSV-M incidents.

8.3.3. Combined Probability

The spreadsheet "Combined" calculates the overall $p(D|I)$ estimates and distribution accounting for all three incident types. To combine the values for VST, TSV-S and TSV-M incidents into one value of $p(D|I)$, the definition of expectation as shown in Equation 3 is again used. In this case, the equation expressing how these factors can be combined is expressed as

$$E[p(D|I)] = \sum_{typacc} f(p(D|I|TYPACC))pr(TYPACC) \quad \text{(Eq. 5)}$$

where TYPACC is the incident type (VST, TSV-S, TSV-M), and $pr(TYPACC)$ is the probability of each incident type occurring based on historical data.

A set of fixed values is used for $pr(TYPACC)$, with different ratios for each warning device type (Table 2). The type of warning device at the crossing appears to influence the type of incident that occurs. Gates appear to significantly reduce the likelihood of VST and TSV-M incidents and most incidents at gated crossings are TSV-S.

Table 8.2. Incident type ratios by warning device type

	Warning Device Type		
Incident Type	Gates	Active	Passive
VST	18.8%	32.1%	26.1%
TSV-S	51.1%	24.4%	23.5%
TSV-M	30.0%	43.5%	50.4%

Beyond the influence of warning device type, it is not clear what factors affect incident type. Whether a train strikes a vehicle or a vehicle strikes a train is likely due to a complex combination of human factors (driver attentiveness, for example) and engineering aspects (reflectorization, crossing lighting, crossing geometry, visibility) at individual crossings. If future researchers develop a more robust understanding of these factors, the calculator could easily be adapted to incorporate a distribution of incident type ratios instead of the fixed values used here.

Note that the definition of expectation in Equation 5 holds when $p(D|I)$ and $pr(TYPACC)$ are independent. This is reasonable based on current understanding of $pr(TYPACC)$, since TYPACC itself is not a factor in the three $p(D|I)$ distributions developed. However, if future research shows there to be common factors, the two will no longer be independent, and this would need to be accounted for by modifying Equation 5.

The spreadsheet "Combined" evaluates Equation 5 by considering the likelihood of each $p(D|I)$ value based on the three TYPACC calculations. This produces the cumulative probability distribution (Figure 8.2). In the same manner, the distribution can be represented by an expected value, which is also provided as an output (Figure 8.1).

Based on the cumulative distribution, the spreadsheet also determines the 95% estimate (and other point estimates) of $p(D|I)$ through interpolation. This means that 95% of the time (based on the distribution of possible conditions during the collision) the likelihood of derailment will be less than the specified value. Due to how the distribution is developed and plotted, it is unlikely that any value of $p(D|I)$ will fall exactly on the 95% line, so the spreadsheet determines the values of $p(D|I)$ just below and above 95%, then interpolates using a linear approximation. The 95% estimate is presented as an output of the calculator (Figure 8.1).

8.4. CALCULATOR USAGE GUIDE

The calculator can be used in two ways. The first is *single crossing mode*, where the user manually enters the characteristics of a crossing on the "Calculator" spreadsheet. This is useful for studying how derailment likelihood would change if crossing conditions were altered.

The second is *batch processing mode*. The user provides records for multiple crossings and activates a macro to automatically calculate the derailment likelihood for each crossing. This is useful for evaluating derailment likelihood for crossings in a corridor or state. To use this mode, the user goes to the "Data" spreadsheet and provides the requested information starting in the second row of the first five columns – warning device type (WDTYPE), highway classification (HWYCLASS), number of passenger trains per day (PASSCNT), total number of trains per day (TRNCNT), and track class (TRKCLAS). The values of WDTYPE and HWYCLASS must match the list of possible values shown in Appendix E. TRKCLASS must be a number between 1 and 5 (provided as a numeral). PASSCNT and TRNCNT must also be numeric, and TRNCNT must be greater than zero. The user may also provide a grade crossing identification number (GXID) that is not used in the calculation of $p(D|I)$ but assists in identification of the crossing for subsequent analysis.

Once information has been supplied on the "Data" spreadsheet, the user presses the "Batch Run" button on the "Calculator" spreadsheet to activate the batch processing macro. The macro enters each record into the calculator, and returns the distribution, expected value and point values of the derailment likelihood for each crossing. The user sorts the final data by any column using the filter arrows.

8.5. DISCUSSION

In previous chapters, the performance of the model and its components was evaluated in multiple ways. Evaluating the performance of the calculator specifically is more complex, since it relies not only on the underlying conditional derailment probability model but also the proxy variable distributions. Given that the calculator inputs are used to select appropriate proxy distributions, instead of directly serving as physical model variables, it is important to understand the effect of each calculator input variable on the estimate of derailment probability. To examine this, the calculator's sensitivity to each input factor was evaluated using a tornado diagram.

While the calculator has five input factors (warning device type, highway class, track class, number of freight trains and number of passenger trains), in practice the calculator combines the number of freight and passenger of trains into a single variable quantifying the percent passenger traffic. Therefore, the sensitivity analysis will use percent passenger traffic as a factor, along with warning device type, highway class and track class. The range and midpoint were identified for each factor (Table 8.3). For each categorical factor, the low, mid and high categories were identified by evaluating the calculator at each factor level and determining which had the smallest and largest resulting expected $p(D|I)$ values.

Table 8.3. Factor level definitions

Factor	Low	Mid	High
% Passenger	0%	50%	100%
Track Class	1	3	5
Highway Class	Urban Local	Rural Local	Rural Arterial
Warning Device Type	Passive	Other Active	Gates

To perform the sensitivity analysis, the calculator was first used to evaluate the derailment model with all factors at the mid level. This yielded the baseline expected *p(D|I)*, 0.854. Subsequently, for one factor at a time, each factor was evaluated at the low and high levels, providing 8 values. This demonstrates the range of *p(D|I)* values based on the range of model factors (Table 8.4, Figure 8.6). Of the three categorical variables, variations in track class have the greatest effect on predicted derailment likelihood, followed by highway class and warning device type.

Table 8.4. Calculator results for each factor level

Factor	Low	High
% Passenger	0.0407	0.1764
Track Class	0.0475	0.1218
Highway Class	0.0584	0.1072
Warning Device Type	0.0804	0.0942

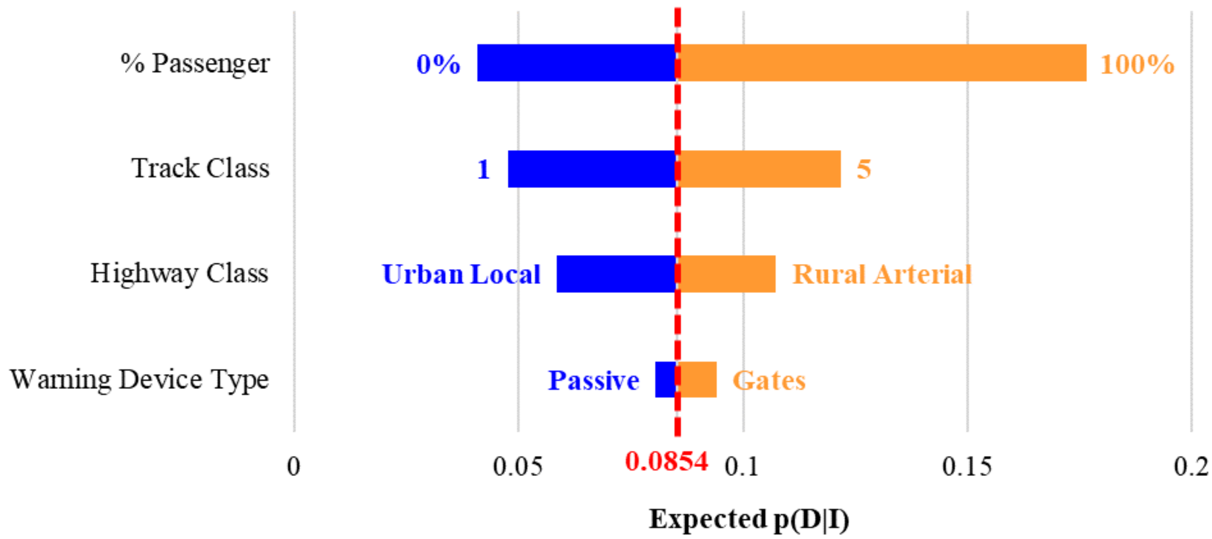


Figure 8.6: Tornado diagram depicting variance of expected $p(D|I)$ created by varying each model factor.

It is more difficult to compare the effect of the continuous variable, percent passenger trains, to the three categorical variables. Since the categorical variables were examined over their entire high-to-low range, I initially also examined percent passenger traffic over its whole range (Figure 8.6). In this case, percent passenger traffic has a greater effect on the expected *p(D|I)* than the other three variables, which is logical given the strong difference in derailment likelihood between freight and passenger trains (Chapter 6). However, sensitivity analyses more typically examine the effect on the result of a calculation based on small variances in the factor values – typically 5-10%. This has a strong effect on the apparent influence of each factor on calculator results (Figure 8.7). If a smaller range of percent passenger traffic is considered, then it has a similar effect on the overall *p(D|I)* calculation as warning device type. This is a significant difference compared to the first analysis.

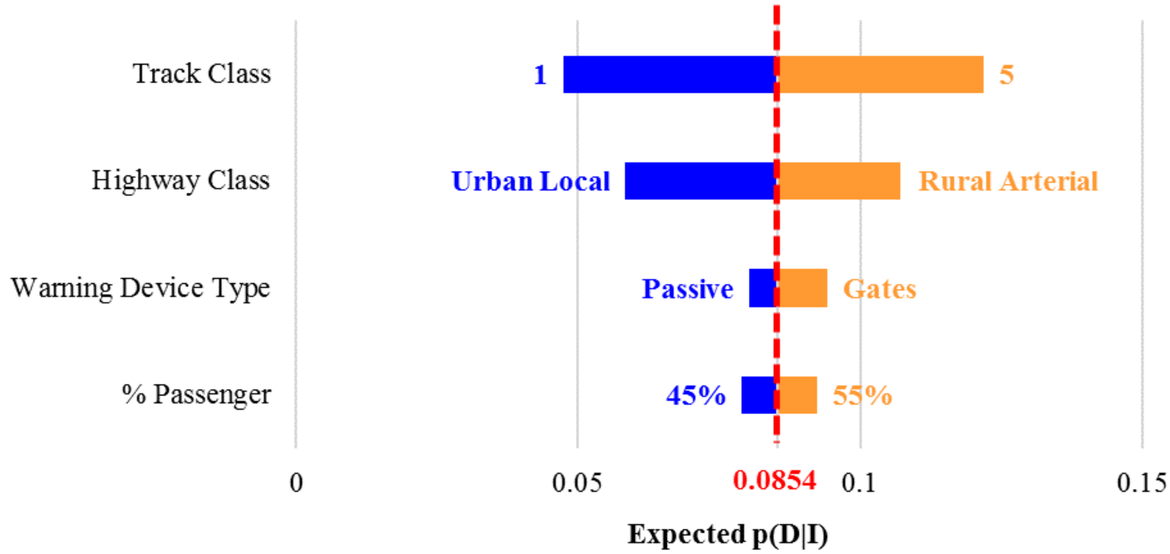


Figure 8.7: Tornado diagram depicting variance of expected $p(D|I)$ created by varying each model factor, but with percent passenger traffic varying from 45% to 55% instead of 0% to 100%.

Tornado diagrams as a form of sensitivity analysis are a useful way to envision model variability, but due to the nature of the calculator, which mostly relies on categorical variables, it may not be as useful in this instance. The tornado diagram is just one way to consider a model's sensitivity to factor variance; there are other potential ways to examine variance, including propagating the error forward at each step to determine overall error in the calculator.

8.6. CONCLUSION

This chapter explained the development of a derailment likelihood calculator that combines the statistical models and distributions developed in Chapters 6 and 7 using Microsoft Excel. The user provides five variables describing conditions at the crossing, and the calculator returns the distribution and point values of $p(D|I)$, the conditional probability of derailment.

Chapter 9 will present a case study demonstrating the results of the derailment likelihood calculator on a sample corridor.

CHAPTER 9: EVALUATION OF THE DERAILMENT LIKELIHOOD CALCULATOR USING A CASE STUDY CORRIDOR OF HIGHWAY-RAIL GRADE CROSSINGS

9.1. INTRODUCTION

In Chapter 8 I described a spreadsheet model that calculates the conditional probability of derailment given that a grade crossing collision has occurred. In this chapter I present an example of how the calculator can be used to provide an additional criterion for ranking grade crossings in terms of derailment likelihood by combining the calculator's output with the output of an incident likelihood model. Derailment likelihood has not traditionally been used to prioritize grade crossing warning upgrades or elimination.

On an annual basis, each state is allocated funds from the U.S. DOT for grade crossing upgrades. It is up to each state to decide how to use them. Typically, states use some form of incident prediction model (such as the U.S. DOT Accident Prediction Formula or a state-specific model (FRA, 1987; Ogden and Korve Engineering, 2007; Benekohal and Elzohairy, 2001)) to determine which crossings to upgrade, as well as what warning devices to install during that upgrade. The only formally-considered grade crossing risk components in the current systems are the risk to highway users. As discussed in Chapter 1, this ignores other sources of risk at the crossing, particularly the risk related to derailment occurrence. If a derailment occurs, then not only highway users but also train passengers, train crew and potentially others in the vicinity will be affected. If a hazardous materials release results, the general public can also be affected, as well as the environment. Therefore, understanding the risk of derailment can help quantify these additional risks.

The derailment modeling tool developed in the preceding chapters can be used in multiple ways to enable consideration of derailment likelihood. Its output can be combined with an incident likelihood model, yielding a ranking of crossings by their derailment risk. It can also be used to calculate derailment risk over a rail corridor, considering total risk, average crossing risk or per-mile risk. The derailment likelihood model might also be used as a "tie-breaker" in cases where multiple crossings have similar collision likelihood but differ in their derailment risk. Each of these uses will be demonstrated. In this example, I use the U.S. DOT Accident Prediction Formula to model incident likelihood. I specifically used the FRA Web Accident Prediction System (WBAPS) to determine values for each crossing. The same techniques discussed in this chapter could work with other incident prediction models.

9.2. CASE STUDY CORRIDORS

Use of the grade crossing derailment calculator will be demonstrated using four rail corridors in the state of Illinois. Each corridor is six miles long. These corridors were selected because I have first-hand familiarity with each of them, and could therefore verify the accuracy of the data from the databases I used. I selected corridors that exhibit a variety of characteristics. 63% of the case study crossings have gates, so gates are overrepresented compared to the state-wide ratio of 45%. Additionally, 8% of the study crossings are passive, compared to the state-wide ratio of 31%. Gates are overrepresented because Corridor 1, which has very high train volume, is completely gated. The case study corridors also have different ratios of passenger traffic. The ratio of freight to passenger traffic seems to have the greatest effect on derailment likelihood, since the effect of all factors change depending on the type of train involved.

Corridor 1 is a segment of the BNSF Metra line running from downtown Chicago to the western suburbs. It is in an urban area and has 17 grade crossings on class 4 track, with 160 trains per day, of which between 104 and 112 are Metra passenger trains (depending on the location along the corridor).

Corridor 2 is a segment of the Illinois Central track (owned by Canadian National) running from north to south through Champaign, IL. It has six grade crossings on class 5 track. It has fewer than 30 trains per day, six of which are passenger (specifically Amtrak) trains. It is in a predominantly rural area though it passes through some small urban areas.

Corridor 3 is a segment of the former Elgin, Joliet and Eastern, now owned by Canadian National. Since it runs through the suburbs around Chicago, it is in an urban area. It has five grade crossings on class 3 track, with 12 trains per day and no passenger traffic.

Corridor 4 is a segment of track owned by Norfolk Southern that runs from east to west through Champaign, Illinois. It is in an urban area with 23 crossings mostly on class 1 track (with a small portion of class 3 track). It has 2-4 trains per day and no passenger traffic.

Multiple comparisons can be made of the crossings, both within and between corridors. Considering risk rankings within one corridor shows the effect of warning device and highway classification, since these factors vary with crossing. By contrast, track class and number of trains tend to stay constant over longer railroad segments, and therefore for multiple crossings in a row. As a result, the effect of track class and number of trains is best illustrated by comparing different corridors. Lastly, risk on a corridor will be related to the density of crossings, not just the probability of derailment at individual crossings. This can be expressed in multiple ways: as the expected yearly derailment frequency, the average derailment frequency per crossing, and the average derailment frequency per mile.

9.3. METHODOLOGY

For each of the grade crossings in the case study, the required factors (warning device, highway classification, number of passenger trains, total trains per day, and track class) were used as inputs to the derailment likelihood calculator. This produced the six estimates of derailment likelihood $p(D|I)$ (expected value, 5th percentile, 25th percentile, 50th percentile, 75th percentile, and 95th percentile). For simplicity, this case study compares only the expected value ($p(D|I)_{exp}$), but the same analysis could be done using any of the point estimates. If a different point estimate is used, the ranking would be expected to change, though the subsequent analysis technique would be the same. Those who use the model can decide which point estimate to use, depending on their risk preferences.

Since the derailment likelihood calculator produces a conditional probability, it must be combined with an incident likelihood model to give derailment probability. In this chapter, the calculator is demonstrated using the U.S. DOT Accident Prediction Formula, the most commonly used incident model in the U.S. This model produces an expected yearly incident frequency (not probability) at a grade crossing, $f(I)$. This means that, combined with the $p(D|I)$ values from the calculator, it will produce an expected derailment *frequency*. Other incident prediction models produce expected incident *probability*; if combined with one of these models, the results of the calculator would produce an expected derailment probability.

The values of $f(I)$ for each crossing were calculated using (WBAPS) (FRA, 2013). This web service generates current reports of the 10-year collision history and predicted collisions per year – which includes an adjustment based on the 5-year collision history at the crossing. The calculation is based on the physical characteristics of the crossing as reported by railroads and state transportation departments to the FRA National Inventory File (FRA, 2016). The quality of the data in the inventory is known to vary by state, as some states are better than others about updating information. Therefore, all data for the case study were selected from the state of Illinois, which is generally considered to have reliable data.

Once $p(D|I)$ and $f(I)$ are determined for each crossing, their product is calculated to give the expected number of derailments at the crossing per year, $f(D)$. Here, two values are presented: the product of $f(I)$ with the expected value of $p(D|I)$ ($f(D)_{exp}$), and the product of $f(I)$ with the 95% confidence estimate of $p(D|I)$ ($f(D)_{95}$)

9.4. RESULTS

The results in this section are presented in tabular form. Tables 9.1, 9.2, 9.3, and 9.4 show an assigned crossing name, the crossing's characteristics, and the values of the five incident probability or likelihood estimates: $p(D|I)_{exp}$, $p(D|I)_{95}$, $f(I)$, $f(D)_{exp}$, and $f(D)_{95}$. For each estimate, a numerical rank is also assigned. A rank of one (1) means the crossing has the highest value of an incident estimate for that corridor, while higher ranks have lower incident estimates. Broadly, the "worst" or "most dangerous" crossing will have rank one.

A red-orange-green color gradient is applied to each rank column to assist readers in detecting how crossings vary in the five estimates. Red is assigned to rank one. The less red (or more green) a cell, the less likely that crossing is to experience an incident. Each table is sorted by largest to smallest value of $p(D|I)_{exp}$, meaning from most to least likely to derail, given that an incident has occurred.

Table 9.1: Ranking of derailment occurrence for Corridor 1 grade crossings

Crossing	Warning Device	Highway Class	Passenger Trains	All Trains	Timetable Speed	Track Class	Expected Value		95th Percentile		WBAPS Prediction		p
							$p(D\|I)_{exp}$	Rank	$p(D\|I)_{95}$	Rank	f(I)	Rank	f(D
1F	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.041411	13	0.00
1C	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.038979	14	0.00
1I	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.036779	15	0.00
1D	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.034671	16	0.00
1G	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.031401	17	0.00
1A	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.109700	4	0.02
1H	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.099278	5	0.02
1B	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.055618	6	0.01
1K	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.053709	8	0.01
1J	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.047705	10	0.01
1E	Gates	UA	112	160	70	4	0.224791	3	0.4777419	2	0.217629	1	0.04
1O	Gates	UC	104	160	70	4	0.206134	4	0.4728943	5	0.051150	9	0.01
1N	Gates	UC	104	160	70	4	0.206134	4	0.4728943	5	0.044397	11	0.00
1M	Gates	UC	104	160	70	4	0.206134	4	0.4728943	5	0.042748	12	0.00

Table 9.2: Ranking of derailment occurrence for Corridor 2 grade crossings

Crossing	Warning Device	Highway Class	Passenger Trains	All Trains	Timetable Speed	Track Class	Expected Value		95th Percentile		WBAPS Prediction		
							$p(D\|I)_{exp}$	Rank	$p(D\|I)_{95}$	Rank	f(I)	Rank	f(D
2E	Gates	RC	6	30	79	5	0.094385	1	0.392332	1	0.016437	2	0.00
2A	Gates	UL	6	24	79	5	0.087156	2	0.339894	5	0.002652	6	0.00
2F	Gates	RL	6	30	79	5	0.084577	3	0.355893	3	0.010876	5	0.00
2C	Gates	UC	6	24	79	5	0.084418	4	0.338246	6	0.015562	4	0.00
2B	Passive	RL	6	24	79	5	0.076929	5	0.356790	2	0.041052	1	0.00

Table 9.3: Ranking of derailment occurrence for Corridor 3 grade crossings

Crossing	Warning Device	Highway Class	Passenger Trains	All Trains	Timetable Speed	Track Class	Expected Value		95th Percentile		WBAPS Prediction		f
							$p(D\|I)_{exp}$	Rank	$p(D\|I)_{95}$	Rank	f(I)	Rank	
3D	Gates	UL	0	11	45	3	0.035933	1	0.212708	1	0.016985	4	0.
3E	Gates	UL	0	11	30	3	0.035933	1	0.212708	1	0.013576	5	0.
3C	Gates	UA	0	11	30	3	0.035725	2	0.212279	2	0.022776	2	0.
3B	Gates	UA	0	11	45	3	0.035725	2	0.212279	2	0.018099	3	0.

Table 9.4: Ranking of derailment occurrence for Corridor 4 grade crossings

Crossing	Warning Device	Highway Class	Passenger Trains	All Trains	Timetable Speed	Track Class	Expected Value		95th Percentile		WBAPS Prediction		f
							$p(D\|I)_{exp}$	Rank	$p(D\|I)_{95}$	Rank	f(I)	Rank	
4U	Other Active	UA	0	2	30	3	0.036895	1	0.224421	1	0.021098	4	0.
4V	Passive	UL	0	2	30	3	0.033726	2	0.191401	4	0.002522	22	0.
4W	Passive	UL	0	2	30	3	0.033726	2	0.191401	4	0.048357	2	0.
4T	Other Active	UC	0	2	30	3	0.033482	3	0.194156	3	0.008774	16	0.
4K	Other Active	UA	0	4	10	1	0.025705	4	0.198818	2	0.012931	9	0.
4N	Other Active	UA	0	4	10	1	0.025705	4	0.198818	2	0.021490	3	0.
4R	Other Active	UA	0	4	10	1	0.025705	4	0.198818	2	0.020922	5	0.
4S	Other Active	UA	0	4	10	1	0.025705	4	0.198818	2	0.058265	1	0.
4C	Other Active	UC	0	4	10	1	0.022299	5	0.160905	5	0.006713	19	0.
4E	Other Active	UC	0	6	10	1	0.022299	5	0.160905	5	0.013107	8	0.
4F	Other Active	UC	0	4	10	1	0.022299	5	0.160905	5	0.008976	15	0.
4H	Other Active	UC	0	4	10	1	0.022299	5	0.160905	5	0.010105	12	0.
4I	Other Active	UC	0	4	10	1	0.022299	5	0.160905	5	0.009618	14	0.
4O	Other Active	UC	0	4	10	1	0.022299	5	0.160905	5	0.012703	10	0.
4P	Other Active	UC	0	4	10	1	0.022299	5	0.160905	5	0.012007	11	0.
4A	Other Active	UL	0	4	10	1	0.021542	6	0.155177	6	0.010027	13	0.
4B	Other Active	UL	0	4	10	1	0.021542	6	0.155177	6	0.014397	7	0.
4G	Other Active	UL	0	4	10	1	0.021542	6	0.155177	6	0.007728	18	0.

9.4.1. Corridor 1

Corridor 1 is in an urban environment. All crossings have gates, as well as the same number of trains and track class (Table 9.1). Therefore, they have similar estimates of *p(D|I)*. There are two passenger train counts: 112 (70% passenger traffic) and 104 (65% passenger traffic). The crossings with 112 passenger trains have larger *p(D|I)* values than those with 104. Of the crossings with 112 passenger trains, crossings 1F, 1C, 1I, 1D, and 1G have the highest value of *p(D|I)*. These are the crossings located on urban local roads. The next group of crossings (1A, 1H, 1B, 1K, and 1J) are on urban connectors, followed by crossing 1E on an urban arterial.

This ordering may appear counterintuitive, since derailment probability increases with increased vehicle speed (Chapters 4 and 6), and vehicle speeds are typically highest on urban connectors, followed by urban arterials and urban local roads. However, highway class also affects the ratio of large to small highway vehicles. Looking at the statistical models used in the calculator (Chapter 6), the large vehicle effect outweighs the vehicle speed effect for most vehicle speeds. Based on the proxy data (Chapter 7), at gated crossings, local roads generally have more large vehicles than connectors or arterials. This is true for all passenger train incidents, and for TSV freight train incidents. Interestingly, this is not true for VST freight train incidents, where local roads have fewer large vehicles. Since Corridor 1 is passenger-dominated, the freight VST effect is not as important in the results.

After the crossings with 112 passenger trains come those with 104 passenger trains. At these crossings, passenger trains constitute a smaller percentage of overall traffic. Since passenger trains are more likely to derail than freight trains, crossings with less passenger traffic have a lower likelihood of derailment.

The ranking by WBAPS-predicted incident frequency is different, as is obvious looking at the color gradient for the *f(I)* rank column. The crossing most likely to have an incident is 1E, which ranks third (of five) in *p(D|I)*. At the other end of the spectrum, the five crossings least likely to have a crossing incident (1F, 1C, 1I, 1D, 1G) all ranked first in *p(D|I)*. Crossing 1E is almost 30 times more likely to experience an incident than crossing 1G (the least likely to have an incident).

Considering the rankings based on expected derailment frequency ($f(D)_{exp}$), the ranking is similar to that produced by *f(I)*. There was not much variation in $p(D|I)_{exp}$, but there was a lot of variation in *f(I)*. As a result, when the product of the two was calculated, the rankings based on $f(D)_{exp}$ differed only slightly from the rankings based on *f(I)* prediction alone.

9.4.2. Corridor 2

Corridor 2 has the greatest diversity of crossing conditions, although there are few crossings in the corridor (Table 9.2). The crossings have different passenger train ratios, a mix of rural and urban highway classifications, and a mix of gated and passive crossings. Due to this blend, it is more difficult to determine exactly which factors lead to the resulting values of *p(D|I)*. Generally, the analysis shows that crossings with gates have higher $p(D|I)_{exp}$ values than passive crossings. Again, this is due to the large vehicle ratio, which covaries with warning device as well as highway class. Additionally, crossings with a higher percentage of passenger trains have higher $p(D|I)_{exp}$ values.

The ranking by WBAPS predicted incident frequency is different from the $p(D|I)_{exp}$ ranking. The crossing most likely to have a grade crossing incident is 2B, which ranks fifth (out of six) in expected conditional probability of derailment. However, crossing 2E has the second highest *f(I)* and the highest $p(D|I)_{exp}$.

The rankings by $f(D)_{exp}$ are almost identical to those produced by $f(I)$. The variation in $p(D|I)_{exp}$ was small compared to the variation in $f(I)$, therefore only small shifts in the ranking occur.

9.4.3. Corridor 3

Corridor 3 is also located in an urban environment. All crossings have gates, as well as a consistent number of trains and track class (Table 9.3). The corridor has no passenger train traffic. Consequently, the crossings all have similar estimates of $p(D|I)$. Crossings 3D and 3E have the highest value of $p(D|I)_{exp}$. These crossings are located on urban local roads. Crossings 3C and 3B are on urban arterials. Crossing 3A is on an urban connector. As explained in Section 9.4.2, this order is explained by the proxy variable for highway vehicle size. Since this corridor has only freight trains, the freight VST effect dominates.

The $f(I)$ ranking is almost opposite to the $p(D|I)_{exp}$ ranking. The crossing most likely to have a grade crossing incident is 3A, which ranks last in expected conditional probability of derailment. At the other end of the spectrum, the two crossings least likely to have a crossing incident (3D and 3E) ranked first in conditional probability of derailment.

Looking at the rankings by expected derailment frequency ($f(D)_{exp}$), the ranking is identical to that produced by $f(I)$. Since there was not much variation in $p(D|I)_{exp}$ the values of $f(I)$ dominate.

9.4.4. Corridor 4

Corridor 4 is also in an urban environment, and its crossings have similar train counts and no passenger traffic. However, this corridor shows more variability in warning device and track class, providing a wider range of $p(D|I)$ (Table 9.4).

The crossings with the highest conditional probability of derailment (4U, 4V and 4W) are those on class 3 track. 4U is on an urban arterial and has bells and flashers (but no gates). The next two crossings in the ranking (4V and 4W) are on urban local roads and have passive warning devices only. They have a lower $p(D|I)_{exp}$, again because of the effect of large highway vehicles, as explained in Section 9.4.2. 4D, the crossing with the lowest $p(D|I)_{exp}$, is the only one with gates. Gated crossings have lower average highway vehicle speeds and a lower percentage of large highway vehicle involvement compared to other warning device types.

The $f(I)$ ranking is very different from the $p(D|I)$ ranking, as is obvious looking at the color gradient for the two columns. Crossing 4S is most likely to have an incident, but ranks fourth (out of seven) in expected conditional probability of derailment. The five crossings least likely to have a crossing incident (4V, 4J, 4L, 4C, and 4G) have a wide range of $p(D|I)_{exp}$ values. 4V has the lowest $f(I)$ but ranks second by $p(D|I)_{exp}$, 4C ranks fifth while the others tie for sixth.

Considering the rankings based on expected derailment frequency ($f(D)_{exp}$), the ranking is similar to that produced by $f(I)$. There is, however, more variability than for the other corridors. A good example is crossing 4T. It ranks 16th out of 22 for $f(I)$ and 3rd for $p(D|I)_{exp}$. When combined, it ranks 9th out of 23 for derailment frequency, $f(D)_{exp}$. This comparatively large change in ranking draws attention to 4T and suggests that it should be investigated as a source of risk.

9.4.5. Inter-corridor Comparisons

In addition to comparing crossings within a corridor, inter-corridor comparisons are of interest as well. First, consider the ranking of derailment likelihood for all grade crossings on all four corridors (Table 9.5). Unlike previous tables, this one is sorted in ascending rank order by $f(D)_{exp}$.

Table 9.5: Ranking of derailment occurrence for all grade crossings

Crossing	Warning Device	Highway Class	Passenger Trains	All Trains	Timetable Speed	Track Class	Expected Value		95th Percentile		WBAPS Prediction		
							$p(D\|I)_{exp}$	Rank	$p(D\|I)_{95}$	Rank	f(I)	Rank	f
1E	Gates	UA	112	160	70	4	0.224791	3	0.4777419	2	0.217629	1	0.
1L	Gates	UA	104	160	70	4	0.204965	5	0.475051	4	0.136538	2	0.
1A	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.109700	4	0.
1P	Gates	UA	104	160	70	4	0.204965	5	0.475051	4	0.118150	3	0.
1H	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.099278	5	0.
1B	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.055618	8	0.
1K	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.053709	10	0.
1Q	Gates	UA	104	160	70	4	0.204965	5	0.475051	4	0.054701	9	0.
1J	Gates	UC	112	160	70	4	0.226153	2	0.4756885	3	0.047705	13	0.
1O	Gates	UC	104	160	70	4	0.206134	4	0.4728943	5	0.051150	11	0.
1F	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.041411	16	0.
1N	Gates	UC	104	160	70	4	0.206134	4	0.4728943	5	0.044397	14	0.
1C	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.038979	18	0.
1M	Gates	UC	104	160	70	4	0.206134	4	0.4728943	5	0.042748	15	0.
1I	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.036779	19	0.
1D	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.034671	20	0.
1G	Gates	UL	112	160	70	4	0.231053	1	0.4778145	1	0.031401	21	0.
2B	Passive	RL	6	24	79	5	0.076929	10	0.356790	7	0.041052	17	0.
3A	Gates	UC	0	12	45	3	0.035169	15	0.204174	15	0.070777	6	0.
4W	Passive	UL	0	2	30	3	0.033726	16	0.191401	18	0.048357	12	0.
2E	Gates	RC	6	30	79	5	0.094385	6	0.392332	6	0.016437	29	0.
4S	Other Active	UA	0	4	10	1	0.025705	18	0.198818	16	0.058265	7	0.
2C	Gates	UC	6	24	79	5	0.084418	9	0.338246	11	0.015562	31	0.
2D	Passive	RL	6	30	79	5	0.070956	11	0.342065	9	0.016346	30	0.
2F	Gates	RL	6	30	79	5	0.084577	8	0.355893	8	0.010876	38	0.
3C	Gates	UA	0	11	30	3	0.035725	14	0.212279	14	0.022776	22	0.
4U	Other Active	UA	0	2	30	3	0.036895	12	0.224421	12	0.021098	24	0.
3B	Gates	UA	0	11	45	3	0.035725	14	0.212279	14	0.018099	27	0.
3D	Gates	UL	0	11	45	3	0.035933	13	0.212708	13	0.016985	28	0.
4N	Other Active	UA	0	4	10	1	0.025705	18	0.198818	16	0.021490	23	0.
4R	Other Active	UA	0	4	10	1	0.025705	18	0.198818	16	0.020922	25	0.
3E	Gates	UL	0	11	30	3	0.035933	13	0.212708	13	0.013576	33	0.

Table 9.5 (cont.): Ranking of derailment occurrence for all grade crossings

Crossing	Warning Device	Highway Class	Passenger Trains	All Trains	Timetable Speed	Track Class	Expected Value		95th Percentile		WBAPS Prediction		
							$p(D\|I)_{exp}$	Rank	$p(D\|I)_{95}$	Rank	f(I)	Rank	f
4I	Other Active	UC	0	4	10	1	0.022299	19	0.160905	19	0.009618	41	0.
4Q	Other Active	UL	0	4	10	1	0.021542	20	0.155177	20	0.009618	41	0.
4F	Other Active	UC	0	4	10	1	0.022299	19	0.160905	19	0.008976	42	0.
4M	Other Active	UL	0	4	10	1	0.021542	20	0.155177	20	0.008225	44	0.
4G	Other Active	UL	0	4	10	1	0.021542	20	0.155177	20	0.007728	45	0.
4C	Other Active	UC	0	4	10	1	0.022299	19	0.160905	19	0.006713	46	0.

Considering all crossings, it is obvious that those on Corridor 1 have the highest values of both $p(D|I)_{exp}$ and $f(I)$. This is likely due to two different, but related factors. The values of $p(D|I)_{exp}$ are high because of the high ratio of passenger to freight trains, since passenger trains are more likely to derail. The values of $f(I)$ are high because Corridor 1 has high train volumes, which increases exposure, the product of the number of trains and highway vehicles that operate over a crossing. Exposure is essentially the number of opportunities that exist for a grade crossing collision to occur. This is a critical factor in the incident likelihood model used by WBAPS. In contrast, crossings on Corridor 4 rank low in both metrics, because it has very few trains, all of them freight.

Comparing the $f(I)$ ranking to the $f(D)_{exp}$ ranking, crossings on either extreme (those with the highest or lowest incident likelihood) generally maintain a consistent place in the ranking. However, there are substantial changes in the middle ranges. For example, crossing 2F ranks 38th when only incident likelihood is considered, but jumps to 25th once derailment likelihood is taken into account. This can also be observed for crossings 4T (43rd to 36th) and 2A (49th to 40th).

It is also possible to compare the total incident or derailment frequency between the corridors (Table 9.6). This total is obtained by summing all the values of $f(I)$, $f(D)_{exp}$, or $f(D)_{95}$ for the corridor.

Table 9.6: Total incident/derailment frequency by corridor

		$f(I)$		$f(D)_{exp}$		$f(D)_{95}$	
Corridor	Crossings	Total	Rank	Total	Rank	Total	Rank
1	17	1.21456	1	0.26595	1	0.578007	1
2	6	0.10293	4	0.00833	3	0.036723	3
3	5	0.14221	3	0.00505	4	0.029628	4
4	23	0.34961	2	0.00901	2	0.062993	2

Generally, corridors with more crossings have higher totals, which is to be expected. However, Corridor 1 has a total *f(I)* value about three times greater than that of Corridor 4, even though Corridor 4 has six more crossings (23 compared to 17). This indicates that both the number of crossings and the total derailment frequency should be considered. Therefore, the frequency was normalized by crossing count to produce an average incident frequency for each crossing on the corridor (Table 9.7).

Table 9.7: Average crossing incident/derailment frequency by corridor

		f(I)		$f(D)_{exp}$		$f(D)_{95}$	
Corridor	Crossings	Average	Rank	Average	Rank	Average	Rank
1	17	0.07144	1	0.01564	1	0.034000	1
2	6	0.01715	3	0.00139	2	0.006120	2
3	5	0.02844	2	0.00101	3	0.005926	3
4	23	0.01520	4	0.00039	4	0.002739	4

A disadvantage of using corridor derailment frequency as a crossing average is that it may overlook the benefits of corridor-based crossing closure programs. Some states have taken the approach of closing several (sometimes more than half) of the crossings in a specific corridor, and upgrading warning devices (or grade separating) the remaining crossings. The idea is to direct vehicle traffic to a limited number of crossings, and then allocate greater resources to safety upgrades at fewer crossings. Since total highway vehicle traffic over the whole corridor typically remains the same, exposure and therefore incident likelihood may increase significantly at the remaining crossings, despite the improved warning devices.

Another way to consider corridor derailment frequency is on a per-mile average. It is not useful for this case study, because the selected corridors are all six miles long; however, in practice, it may make more sense because users will have corridors of non-uniform lengths. If this is the case, normalizing corridor derailment frequency by corridor length will provide a more accurate comparison between corridors.

9.5. DISCUSSION

The output of the derailment likelihood calculator can be considered and interpreted in multiple ways, and each has value depending on particular user questions. Within a corridor, the prioritization of crossings for upgrade may change if derailment likelihood is considered in addition to incident likelihood. However, crossings that are considered most likely to have an incident are typically also the most likely to have a derailment. This means that an effective way to reduce derailment occurrence is to reduce incident occurrence, with whatever crossing upgrades that entails.

Comparing between corridors – especially those of similar lengths – can be especially useful for understanding overall crossing risk. A crossing could have an exceptionally high derailment likelihood, but might be the only crossing for miles because of crossing closure programs (for example). In contrast, another corridor could have many crossings each with much lower derailment likelihoods. At first glance, the single high-likelihood crossing could appear to be a greater source of concern, but total derailment likelihood on the second corridor could be higher, contributing to increased risk.

This is related to consideration of average crossing derailment frequency. If only total corridor derailment frequency is compared, the effect of number of crossings might be overlooked. Normalizing the total by number of crossings mitigates this effect. If a corridor has a high total derailment frequency but low average frequency, it may be of less concern. By contrast, a corridor where both total and average derailment frequency are high would likely merit additional mitigation measures.

Corridor-level analysis is also useful for adding in consideration of consequence metrics for the risk model. A derailment on a corridor in an urban, densely-populated area could have much higher consequences than a similar corridor in a rural, sparsely-populated area.

9.6. CONCLUSIONS

The goal of this chapter was to illustrate and explain how the calculator developed in Chapter 8 can provide guidance to users when determining which grade crossings to prioritize for upgrade. The derailment likelihood model can be used in combination with an incident likelihood model such as the U.S. DOT Accident Prediction Model, yielding a ranking of crossings according to their derailment likelihood. The results can be compared within a corridor of crossings, between multiple corridors, or even across a state as a whole. Each of these techniques can provide users with useful information about grade crossing safety.

CHAPTER 10: CONCLUSIONS

10.1. FUTURE RESEARCH NEEDS

Since the topic of grade crossing derailment risk had previously not been researched in depth, there are a number of directions for future research that could improve our understanding of this risk. I developed a robust derailment likelihood model, but as with any model there are areas where it could continue to be further refined. Here, I present some future work that would improve our understanding of factors affecting grade crossing safety.

10.1.1. Data Availability and Potential Improvements

Researchers in the U.S. are fortunate to have a large dataset collected by FRA concerning grade crossing collisions and other railroad incidents. A number of variables are recorded, enabling researchers to address numerous questions with the existing data. FRA is also reactive to new research questions, occasionally adding data fields in response to new safety concerns.

However, as with any data source, there are other factors that would further improve the utility of the databases. I discuss a number of these factors in the pertinent chapters in my dissertation including explanations of both the database usage and some of the limitations in terms of the variables that are present, variable fields provided, and completeness of the data in the existing fields.

In the GCI database, the crossing angle variable should provide exact crossing angle, or at least a more fine-grained categorization of the angle. This would improve the ability to analyze a variety of aspects of the physical factors affecting collisions, including the effect on derailment probability. Additionally, more detail regarding both track and highway curvature and grade, and aspects of crossing geometry including humped crossings, would be beneficial for evaluating the effect on safety and risk of crossing design factors.

For the HRA, more detailed information about the type of highway vehicle involved in collisions, including its weight, and the exact angle of collision between the train and the vehicle would remove the need for many of the assumptions necessary to estimate crossing risk. Currently, the angle of collision can only be determined by combining information from the GCI on the crossing angle category with information from the HRA about the cardinal direction (north, south, east, or west) each vehicle was traveling at the time of the collision, and then comparing this information to Google Earth images of the crossing. This process is very time consuming, and also relies heavily on interpretation of the data. In addition to these data, information on damage to the grade crossing warning system in an incident would improve the ability to determine if the increase in REA-reportable grade crossing incidents is simply due to more costly equipment being present at more crossings, or if something else is occurring.

In the REA database, the type of rail vehicle involved in either a VST or TSV incident should be defined (locomotive, railcar, EMU, etc.), and its weight should be included. This would improve the utility of the data for derailment prediction and other aspects of grade crossing incident safety analysis. The position of each locomotive in the train consist, especially if distributed power is being used, should be defined. More generally, complete consist information in the REA database would enhance the ability to address a variety of rail safety questions, not just at grade crossings. Additionally, many individuals who commented on this work over the course of its development asked if emergency brake application played a role in derailment occurrence. Adding a field indicating if emergency brakes were applied would answer this question.

At a higher level, a helpful addition would be new field with a standard incident number to enable unambiguous linkage of incidents reported to the HRA with the report for the same incident in the REA database (if it met the criteria for inclusion in the latter). To link the data, I developed a technique for creating a unique incident ID relying on the date and location of each incident in the REA and HRA databases. However, this required certain assumptions and additional data processing that could be avoided if a standard incident number was provided. This would bring all REA variables into play for HRA analyses whenever an incident exceeds the REA threshold, and vice versa, substantially leveraging the value of both databases with little additional effort.

Perhaps the greatest value that could be added to all three databases is better enforcement of completion of all fields in a form when it is submitted, as well as cross-checking logic as appropriate to make sure different fields do not contain conflicting data. Due to incomplete, incorrect or ambiguous data, I discarded thousands of records from the analysis that could otherwise have provided useful information and improved the resolution of the results presented here. Such improvements in reporting accuracy and consistency would have broader benefits in terms of improving the utility of these FRA databases that transcend the objectives of my research. The three databases varied in terms of the completeness of the information. The REA and HRA were roughly similar, but both were more complete than the GCI. The quality and completeness of the GCI database varied the most of the three. It is the responsibility of individual states to update the GCI data, and states vary widely in their reliability in updating the GCI as crossing circumstances change over time. Further details of opportunities to improve the databases and their utility to the U.S. highway and railway safety community can be found in the several of the chapters herein.

10.1.2. LIDAR Data Opportunities

Some important opportunities will be available to future researchers in the coming years. FRA is currently collecting locomotive-mounted LIDAR data on grade crossings across the U.S. LIDAR takes a 3-D scan of the area surrounding the measuring device, which in this case is mounted on a locomotive. It can be used to measure clearances (it has been used to assist double-stack operation), check for encroachment of trees or other debris on the track, etc. Of interest to grade crossing research, it could assist with identification of additional variables. For example, many railroad practitioners who commented on the work in this dissertation said they believed (based on their experience) that track superelevation and curvature influenced derailment likelihood. Currently, there is no way to know these factors based on information in the FRA databases (though it could be manually cross-referenced from track charts if desired). It would also be useful for identifying humped crossings from the database, a factor which currently can only be evaluated through visual inspection of crossings. The LIDAR project could also facilitate accurate determination of the crossing angle. Depending on how the LIDAR data are made available, these factors could be determined based on the output.

Even if the LIDAR data are not made available to researchers in their entirety, it could be possible for FRA to include additional fields for these specific factors. The GCI in particular could include these fields by requesting such data from the crossing's host railroad. The forms providing information for the GCI are already being completed by railroad employees (as of 2015 it was a requirement) and it could be possible to provide more detail.

Having data about these additional factors would enable refinement of the models developed here, thereby potentially leading to more accurate evaluation of derailment probability.

10.1.3. Proxy Variables

The proxy variables developed in this dissertation were developed to serve as "average" or "typical" conditions at a grade crossing, since it is not possible to determine the exact distributions (for example, highway vehicle speeds) at specific crossings based on available information. Two approaches could be taken to improve the results of the calculator by improving or removing the need for proxies. First, researchers could conduct field studies to determine if the proxies discussed here are faithful to real-life scenarios, or to develop more accurate proxies. Second, the calculator could be modified so practitioners can provide their own distributions based on field studies conducted as part of a crossing upgrade project. The latter approach would probably provide the most accurate estimation of derailment likelihood at a crossing. A piece of software could be developed that allows users to provide detailed data, or select from the available "preset" data developed with the proxy variables.

10.1.4. Model Adaptability

It would be interesting to know if the models developed in this dissertation, which were developed using U.S. incident data, work for predicting derailments in other countries. The U.S. has substantially more freight than passenger traffic. Compared to many other countries this pattern is unusual. North American rail equipment is also heavier than that of most other countries. This is especially true for passenger rail equipment, which is required to comply with robust crashworthiness standards due to the potential for collisions between passenger and freight rail equipment (a hazard that is less common in most other countries).

Refining the model using international data would not only make the model globally applicable and therefore helpful to more people, but would provide additional predictive ability to the model in the U.S. For example, the model shows that lighter rail equipment is more likely to derail. What would happen to this likelihood if the U.S. were to begin higher speed rail service using lighter passenger rail equipment? Using only U.S. data, it is not possible to answer this question analytically, but incorporating international data could enable such analysis.

10.1.5. Incorporation into Holistic Risk Framework

The ultimate potential of this model depends on how it is incorporated into the overall railroad safety picture. By incorporating collision likelihood models as well as consequence models, it would be possible to fully illustrate grade crossing risk.

10.1.5.1. Collision Modeling

Considerable research has been conducted related to grade crossing collision likelihood modeling. However, there is room to expand on the existing knowledge base and explore new techniques for improving safety, especially in the context of the impacts on railroad safety.

Collision likelihood modeling should continue to develop and evolve. New research may be conducted to understand if the U.S. DOT Accident Prediction Formula needs to be updated to reflect the current state of technology, or if an entirely new formula needs to be developed. The latter approach has been undertaken by a variety of researchers to date, producing a variety of more accurate models. These may be useful in replacing or augmenting the U.S. DOT formula, but since many use small, regional data sets to develop regression models, it should first be understood if these are applicable on a wider scale. Current formulas do not differentiate between four-quad gates or sealed corridor approaches, compared to traditional two-quad gated crossings. As use of these and other augmented warning systems increases, understanding how they affect collision likelihood is important.

10.1.5.2. Consequence Modeling

Accurately representing the expected consequences of a grade crossing collision is another important aspect of quantifying grade crossing risk. Many of the potential consequences have already been investigated to some degree, including hazardous materials releases, highway user casualties, and train passenger casualties. However, more work is required to determine how these models can be incorporated into a holistic grade crossing risk model.

Additional factors that have not been modeled previously – such as train crew casualties – should be considered as well. Data on such factors can be found in the HRA database. Researchers can use these data to develop additional models.

Once the consequences have been modeled, they can be combined with the incident and derailment likelihood models to accurately compare risk between crossings. For example, two identical crossings may have the same incident and derailment likelihoods, but different consequences in the event of a hazardous materials release, based on population density, geography, and other features of the area.

This grade crossing risk model should also take into account how decision makers react to uncertainty – i.e., how their risk aversion might affect the incorporation of various consequences. For example, it is likely that the majority of casualties due to grade crossing collisions are incurred by highway vehicle occupants, and comparatively few affect train passengers or people living near rail lines. However, a single catastrophic grade crossing derailment that results in a large number of casualties to passengers or the general public may be viewed more negatively than a series of smaller incidents where only one or two people are injured or killed.

10.1.5.3. Corridor-Based Risk Assessment

To maximize the utility of this and other railroad risk models, they could be combined into a single, corridor-based risk model. The railroad industry is often faced with choices regarding where it can invest in safety improvements that may reduce risk. Currently, these areas are considered separately; however, if they were all combined into an integrated model, it might be possible to determine the greatest sources of risk on a rail corridor and invest accordingly.

Additionally, incorporating all railroad risks in this matter would help with route planning for hazardous materials and other sensitive cargo. Users could determine the tradeoffs between, for example, a corridor with a large number of grade crossings and higher FRA track class, and one with fewer crossings but lower FRA track class.

10.2. CONCLUSION

This dissertation developed a comprehensive statistical understanding of the factors that affect the occurrence of derailments caused by highway-rail grade crossing collisions. A core set of physical factors were identified and a set of regression models developed that are capable of identifying grade crossing incidents that are more or less likely to result in derailment. These models were then adapted to account for crossing-specific factors, using a set of proxy variables – a series of relationships between incident-specific factors and crossing characteristics. These proxy variable relationships were developed based on the study data and provide "expected" relationships based on average crossing conditions, a necessary provision in the absence of crossing-specific relationships.

The regression models and proxy variables were incorporated into an Excel calculator that prompts users for selected crossing characteristics (highway classification, warning device, timetable speed, number of trains per day and number of passenger trains per day). The model equations are then evaluated based on the distributions of other factors generated by these crossing characteristics. The calculator provides an expected value of the conditional probability of derailment ($p(D|I)$) as well as several point estimates based on percentile. An additional output is the $p(D|I)$ distribution, that shows how likely each value of $p(D|I)$ is to occur, based on the likelihood of different factor combinations.

I concluded by explaining how the calculator can be used to consider derailment likelihood in a case study. Incident likelihood (and the associated risk to highway users) is just one aspect of grade crossing risk. Derailment likelihood, another aspect of that risk, must be considered if we are looking at the overall safety implications of grade crossings. My research adds this new dimension to our understanding of how to assess grade crossing risk and warning system upgrade prioritization. The model allows users to identify crossings with high derailment likelihood, something that was not previously possible. This model will enable more informed allocation of safety resources to minimize risk due to grade crossings. It can be integrated into an overarching risk analysis framework that would consider all sources of risk at a grade crossing. Ultimately, such a tool would open up new opportunities for railroad risk reduction, leading to a safer operating environment for railroads, rail passengers, highway users, and the general public alike.

REFERENCES

49 CFR 213.9 (October 1, 2011). *Classes of track: operating speed limits*.

Aldrich, M. 2008. *Death Rode the Rails: American Railroad Accidents and Safety, 1828-1965*. JHU Press.

Anderson, R.T. 2005. Quantitative Analysis of Factors Affecting Railroad Accident Probability and Severity. Master's Thesis, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering. Urbana, IL, USA.

Associated Press. 2015. Passenger Train Collides With Truck in North Carolina. *The New York Times*. URL http://www.nytimes.com/2015/03/10/us/passenger-train-collides-with-truck-in-north-carolina.html?_r=1 . Accessed 2016-03-15.

Association of American Railroads (AAR). 2017. *Freight Rail's Economic Impact*. URL https://www.aar.org/todays-railroads/our-network. Accessed 2017-03-02.

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

Barkan, C.P.L., C.T. Dick and R.T. Anderson. 2003. Railroad derailment factors affecting hazardous materials transportation risk. *Transportation Research Record: Journal of the Transportation Research Board.* 1825: 64-74.

Benekohal, R.F. and Y.M. Elzohairy. 2001. A new formula for prioritizing railroad crossings for safety improvement. In: *Proceedings of the Institution of Transportation Engineers Annual Meeting*, Chicago, IL.

Bien-Aime, P., 2009. *North Carolina "Sealed Corridor" Phase I, II and III Assessment*. Federal Railroad Administration Office of Research and Development, Washington, DC.

Brod, D. et al., 2013. *Comprehensive Costs of Highway-Rail Grade Crossing Crashes*. NCHRP Report 755. Transportation Research Board, Washington, DC.

Caird, J.K., Transportation Development Centre (Canada), University of Calgary and Cognitive Ergonomics Research Laboratory, 2002. *A Human Factors Analysis of Highway-Railway Grade Crossing Accidents in Canada.* Transportation Development Centre, Montreal.

Cairney, P., 2003. *Prospects of improving the conspicuity of trains at passive railway crossings, CR 217.* Australian Transport Safety Bureau, Canberra.

Carroll, A.A. and M. Haines. 2002a. *North Carolina "sealed corridor" phase I safety assessment.* Transportation Safety Board [CD-ROM]. Transportation Research Board, Washington, DC.

Carroll, A.A. and M. Haines. 2002b. *The use of photo enforcement at highway-rail grade crossings in the U.S.* Transportation Safety Board [CD-ROM]. Transportation Research Board, Washington, DC.

Chadwick, S.G., M.R. Saat and C.P.L. Barkan. 2012. Analysis of Factors Affecting Train Derailments at Highway-Rail Grade Crossings. In: *Proceedings of the Transportation Research Board 91st Annual Meeting*, Washington, DC, USA.

Chadwick, S.G., N. Zhou, and M.R. Saat. 2014. Highway-rail grade crossing safety challenges for shared operations of high-speed passenger and heavy freight rail in the U.S. *Safety Science* 68, 128-137.

Chaudhary, M., A. Hellman, and T. Ngamdung. 2011. *Railroad right-of-way incident analysis research.* Federal Railroad Administration Office of Railroad Policy and Development, Washington, DC.

Cherchas, D.B., G.W. English, E.R. McIlveen, and C. Schwier. 1979. *Development of a Digital Computer Program for Grade Crossing Collision Analysis*. CIGGT Report No. 79-14. Canadian Institute of Guided Ground Transport, Queen's University at Kingston, Ontario.

Cherchas, D.B., G.W. English, N. Ritchie, E.R. McIlveen, and C. Schwier. 1982. Prediction of the Probability of Rail Vehicle Derailment During Grade Crossing Collisions. *Journal of Dynamic Systems, Measurement, and Control* 104, 119-132.

Code of Federal Regulations (CFR), 2012. "Transportation." Title 49 Code of Federal Regulations, Part 213-222.

Code of Federal Regulations (CFR), 2012. "Railroad Accidents/Incidents: Reports Classification and Investigations." 49 CFR, Part 225.

Dick, C.T. 2001. *Factors Affecting the Frequency and Location of Broken Railway Rails and Broken Rail Derailments.* Master's Thesis, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering. Urbana, IL, USA.

Enoch, H. 2014. *The Red-Eyed Guard: A History of Grade Crossing Signs and Signals*. URL http://www.friendsoftheflange.com/2014/10/for-most-people-railroad-is-in-literal.html. Accessed 2017-03-19.

Evans, L. 1991. *Traffic Safety and the Driver*. Van Nostrand Reinhold, New York.

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

Federal Highway Administration (FHWA), 2012. *Manual on Uniform Traffic Control Devices for Streets and Highways, 2009 Edition*. URL https://mutcd.fhwa.dot.gov/pdfs/2009r1r2/mutcd2009r1r2edition.pdf. Accessed 2017-03-10.

Federal Railroad Administration (FRA), 1987. *Rail-Highway Crossing Resource Allocation Procedure: User's Guide, 3rd Edition.* Office of Safety Analysis, Washington, DC.

Federal Railroad Administration (FRA), 2006. *Report to the House and Senate Appropriations Committees: The Safety of Push-Pull and Multiple-Unit Locomotive Passenger Rail Operations.* Office of Railroad Development, Washington, DC.

Federal Railroad Administration (FRA), 2009a. *Highway-Rail Grade Crossing Guidelines for High-Speed Passenger Rail.* Office of Railroad Development, Washington, DC.

Federal Railroad Administration (FRA), 2009b. North Carolina Department of Transportation's "Sealed Corridor" Assessment – Phase IV. In: *Research Results*, Washington, DC.

Federal Railroad Administration (FRA), 2009c. *High-Speed Passenger Rail Safety Strategy*. Office of Railroad Development, Washington, DC.

Federal Railroad Administration (FRA), 2010. Use of Traffic Channelization Devices at Highway-Rail Grade Crossings. In: *Research Results*, Washington, DC.

Federal Railroad Administration (FRA), 2011a. *Download Accident Data On Demand*. URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx. Accessed 2011-03-15.

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

Federal Railroad Administration (FRA), 2011c. Low-Cost Warning Device Industry Assessment. In: *Research Results*, Washington, DC.

Federal Railroad Administration (FRA), 2012a. *Crossing Inventory by State*. URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/invtab.aspx. Accessed 2012-01-15.

Federal Railroad Administration (FRA), 2012b. *Highway/Rail Incidents Summary Tables: Jan 2011 - Dec 2011.* URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/gxrtab.aspx. Accessed 2012-5-14.

Federal Railroad Administration (FRA), 2013. *FRA Web Accident Prediction System (WBAPS),* FRA, U.S. Department of Transportation, 2013.

Federal Railroad Administration (FRA), 2014. *FRA Office of Safety Analysis Web Site: 9.06 Railroad Reporting Threshold*. URL safetydata.fra.dot.gov/officeofsafety/default.aspx. Accessed 2015-2-07.

Federal Railroad Administration (FRA), 2015. *One Year Accident/Incident Overview – Combined*. URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/AccidentByRegionStateCounty.aspx. Accessed 2015-03-20.

Federal Railroad Administration (FRA), 2016. *Highway-Rail Crossing Inventory Data*. URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/downloaddbf.aspx. Accessed 2016-12-20.

Federal Railroad Administration (FRA), 2017a. *2.08 Highway-rail Crossings.* URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/gxrtally1.aspx. Accessed 2017-01-18.

Federal Railroad Administration (FRA) 2017b. *National Grade Crossing Inventory Reporting Regulations.* URL https://www.fra.dot.gov/Page/P0856. Accessed 2017-03-15.

Federal Railroad Administration (FRA) 2017c. *4.08 Casualty Summary Tables.* URL http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/castab.aspx. Accessed 2017-03-16.

Glickman, T.S. and E. Erkut. 2007. Assessment of hazardous material risks for rail yard safety. *Safety Science* 45 (2007) 813-822.

Glover, J. 2009. The Level Crossing Conundrum. *Rail Professional* (November), pp. 30-31.

Hall, S. 2007. Reducing Risk at Automatically-Operated Level Crossings on Public Roads. In: *IET Seminar on Reducing Risk at the Road Rail Interface*, pp. 1-5, 17.

Hauer, E. and B.N. Persaud. 1986. Rail-highway grade crossings: Their safety and the effect of warning devices. In: *Proceedings of the 30th Annual American Association for Automotive Medicine*, Montreal, QC, pp. 247–262.

Hellman, A.D. and A.A. Carroll. 2002. Preliminary evaluation of the school street four-quadrant gate highway-railroad grade crossing. *Transportation Research Board* [CD-ROM], Washington, DC.

Hellman, A.D. and T. Ngamdung. 2009. *Illinois High-Speed Rail Four-Quadrant Gate Reliability Assessment*. Federal Railroad Administration Office of Research and Development, Washington, DC.

Hellman, A.D. and T. Ngamdung. 2010. *Low-Cost Warning Device Industry Assessment.* Federal Railroad Administration Office of Research and Development, Washington, DC.

Hosmer, D.W. and S. Lemeshow. 2002. *Applied Logistic Regression (Second Edition).* Wiley-Interscience Publication.

Jacobsen, K., D. Tyrell, and B. Perlman. 2003. Rail Car Impact Tests with Steel Coil: Collision Dynamics. In: *Proceedings of ASME/IEEE Joint Railroad Conference*.

Jeng, O.-J. 2005. *Survey of Driver Perceptions of Railroad and Light Rail Warning Devices/Grade Crossings*. Federal Transit Administration, Washington, DC.

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

Lenné, M.G., C.M. Rudin-Brown, J. Navarro, J. Edquist, M. Trotter, and N. Tomasevic. 2011. Driver behaviour at rail level crossings: Responses to flashing lights, traffic signals and stop signs in simulated rural driving. *Applied Ergonomics.* 42 (4), 548-554.

Liu, X., M.R. Saat and C.P.L. Barkan. 2012. Analysis of causes of major train derailment and their effect on accident rates. *Transportation Research Record: Journal of the Transportation Research Board.* 2289: 154-163.

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

Modarres, M., M. Kaminskiy, and V. Krivtsov. 2009. *Reliability Engineering and Risk Analysis: A Practical Guide.* 2nd ed. CRC Press, Boca Raton, FL.

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

Morrissey, J. 1980. *The Effectiveness of Flashing Lights and Flashing Lights with Gates in Reducing Accident Frequency at Public Rail-Highway Crossings 1975-1978: Final Report.* Federal Railroad Administration, Washington, DC.

National Cooperative Highway Research Program (NCHRP). 2014. *Fiscal Year 2015 Announcement of Research Projects.*

National Transportation Safety Board (NTSB). 1998. *Safety study: Safety at passive grade crossings, Volume 1: Analysis (PB98-917004, NTSB/SS-98/02).* NTSB, Washington, DC.

National Transportation Safety Board (NTSB). 2014. *Highway Accident Report: Highway-Railroad Grade Crossing Collision: Rosedale, Maryland.* URL https://app.ntsb.gov/doclib/reports/2014/HAR1402.pdf. Accessed 2016-03-15.

National Transportation Safety Board (NTSB). 2015a. *Crash Summary Factual Report: Valhalla, NY.* URL http://dms.ntsb.gov/public/58000-58499/58036/583617.pdf. Accessed 2016-03-15.

National Transportation Safety Board (NTSB). 2015b. *Preliminary Report: HWY15MH006 (Grade Crossing Collision in Oxnard, California).* URL http://www.ntsb.gov/investigations/AccidentReports/Reports/HWY15MH006_preliminary.pdf. Accessed 2016-03-15.

Nelson, A. 2010. Breaking down the barriers to safer crossings. *International Railway Journal* 50 (5), 45-47.

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

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

Raub, R. 2007. Examination of Vehicle Crashes Nationally at Highway-Rail Grade Crossings. *ITE Journal*.

Saat, M.R., and C.P.L. Barkan. 2005. Release Risk and Optimization of Railroad Tank Car Safety Design. *Transportation Research Record: Journal of the Transportation Research Board*, 1916: 78-87.

Saccomanno, F. F., L. Fu, and L.F. Miranda-Moreno. 2004. Risk-Based Model for Identifying Highway-Rail Grade Crossing Blackspots. *Transportation Research Record: Journal of the Transportation Research Board*, 1862: 127-135.

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

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

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

Savage, I. 2006. Does public education improve rail-highway crossing safety? *Accident Analysis and Prevention* 38 (2), 310-316.

Schulte, W. R. 1975. Effectiveness of automatic warning devices in reducing accidents at grade crossings. *Transportation Research Record: Journal of the Transportation Research Board*, 611: 49–57.

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

Simons, J.W., and S.W. Kirkpatrick. 1998. High-Speed Passenger Train Crashworthiness and Occupant Survivability. In: *Proceedings of the International Crashworthiness Conference*.

Stephens, P., and K.M. Foy. 2008. Using Software Tools to Provide Improved Hazmat Visibility for Freight Railroads. In: *Proceedings of the 15th World Congress on Intelligent Transport Systems.*

Sussman, E.D., and T.G. Raslear. 2007. Railroad Human Factors. *Reviews of Human Factors and Ergonomics* 3 (1) 148-189.

Taylor, K. 2008. Addressing road user behavioural changes at railway level crossings. In: *Proceedings of Joint ACRS-Travelsafe National Conference*, 368-375.

Tey L.S., L. Ferreira, and A. Wallace. 2011a. Measuring driver responses at railway level crossings. *Accident Analysis & Prevention*, 43 (6) 2134-2141.

Tey, L.S., G. Wallis, and L. Ferreira. 2011b. Using a driving simulator to assess driver compliance at railway level crossings. In: *Proceedings of the 34th Australasian Transport Research Forum*, Adelaide, Australia.

Transportation Research Board (TRB). 1998. *Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits*. Special Report 254. National Academy Press, Washington, D.C.

Tyrell, D., and A.B. Perlman. 2003. Evaluation of rail passenger equipment crashworthiness strategies. *Transportation Research Record: Journal of the Transportation Research Board*, 1825: 8-14.

Tyrell, D., E. Martinez, K. Severson, K. Jacobsen, D. Parent, M. Priante, and A.B. Perlman. 2006. Overview of a Crash Energy Management Specification for Passenger Rail Equipment. In: *Proceedings of the Joint Rail Conference*.

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

Verma, M., and V. Verter. 2007. Railroad transportation of dangerous goods: Population exposure to airborne toxins. *Computers & Operations Research*, 34(5): 1287-1303.

Walker, F.W., and S.E. Roberts. 1975. Influence of lighting on accident frequency at highway intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 562: 73-78.

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

Wullems C. 2011. Towards the adoption of low-cost rail level crossing warning devices in regional areas of Australia: A review of current technologies and reliability issues. *Safety Science* 49 (8-9): 1059-1073.

Yeh, M., J. Multer. 2008. *Driver Behavior at Highway-Railroad Grade Crossings: A Literature Review from 1990-2006.* Federal Railroad Administration Office of Research and Development, Washington, DC.

Zhang, L. 2017. Unpublished data. Personal communication.

APPENDIX A: DATA CLEAN-UP PROCEDURE

VEHSPD and TRNSPD: If no speeds were reported for a given incident, that incident was omitted from the data set.

LGVEH: The HRA database has a field named TYPVEH. This field defines 11 categories of highway user, including automobiles, semi-tractor-trailers, buses, motorcycles and pedestrians. For the purposes of this study, incidents involving straight trucks and tractor-semitrailers (categories B and C) were defined as "large vehicles" and all others were defined as "small vehicles". All incidents were omitted which were classified as "other motor vehicles," "pedestrian" or "other". The "other motor vehicles" and "other" categories were omitted because it is not possible to reliably identify whether the vehicle is large or small. According to the narrative fields for the "other motor vehicle" entries, the vehicles range from small vehicles such as snowmobiles and SUVs to very large vehicles such as road graders and farm equipment. The "other" category should be used to describe any vehicle involved in a collision which is not a motorized vehicle; however, the narrative field indicates that many of these incidents involved motorized vehicles such as all-terrain vehicles, SUVs, and farm equipment. Other collisions involve bicycles and horse-and-buggies. Since the "other" category is such a mixture of vehicle types, these incidents were omitted. Removing these categories decreased the size of the dataset by about 2,500 entries. It is possible that about half of these could be added back to the dataset if the narrative fields were used to recode the vehicle type manually. However, the dataset is large enough that these entries are unnecessary.

TRNSTK: The HRA database has a field named TYPACC, which indicates whether the train was struck by the highway vehicle, or whether the train struck the highway vehicle. If

TYPACC = 1, then rail equipment struck the highway user, therefore TRNSTK = Y; if TYPACC = 2, then the rail equipment was struck by the highway user, therefore TRNSTK = N.

TRKCLAS, WARNSIG, VIEW and *PUBLIC*: If any of these fields had no value reported, then that data point was excluded.

XTYPE: The HRA database contains a field called CROSSING which lists all of the crossing warning devices in use at the time of the incident. CROSSING has 12 values indicating the type of device. For the purposes of this study, the data were re-categorized into 5 groups, as shown in Table A.1. Any incident record missing the CROSSING data was omitted from analysis.

Table A.1 XTYPE and CROSSING categories

XTYPE	CROSSING
1: Gates	01: Gates
2: Active (excluding gates)	02: Cantilever flashers; 03: Standard flashers; 04: Wig wags; 05: Highway traffic signals; 06: Audible
3: Passive	07: Crossbucks; 08: Stop signs
4: Other	09: Watchman; 10: Flagged by crew; 11: Other (specify)
5: None	12: None

APPENDIX B: TYPEQ CATEGORIES AND PASSENGER/FREIGHT CLASSIFICATION

Table B.1: TYPEQ and Passenger/Freight Classification

Category Number	Category Definition	Number of Derailments	Number of Incidents	Freight/Passenger Classification
1	Freight Train	258	36,727	Freight
2	Passenger Train	38	3,380	Passenger
3	Commuter Train	4	671	Passenger
4	Work Train	7	740	Freight
5	Single Car	0	47	Freight (cars)
6	Cut of Cars	1	45	Freight (cars)
7	Yard or Switching Train	3	3,734	Freight
8	Light Locos	6	2,983	Freight (locos)
9	Maintenance/Inspection Car	10	528	Passenger (cars)
A	Maintenance-of-Way Equipment	1	127	Freight

APPENDIX C: WARNING DEVICE GROUPS

Table C.1: Warning Device Groups

HRA Data File Structure Value	WDCODE	HRA Data File Structure Definition	Chadwick Grouping
01	8	Gates	Gates
02	7	Cantilever Flashing Lights	Active
03	7	Standard Flashing Lights	Active
04	6	Wig Wags	Active
05	6	Highway Traffic Signals	Active
06	6	Audible	Active
07	3	Cross Bucks	Passive
08	4	Stop Signs	Passive
09	1	Watchman	Other (Omitted)
10	1	Flagged by Crew	Other (Omitted)
11	2	Other (Specify)	Other (Omitted)
12	1	None	None (Omitted)

APPENDIX D: SPEED DISTRIBUTIONS

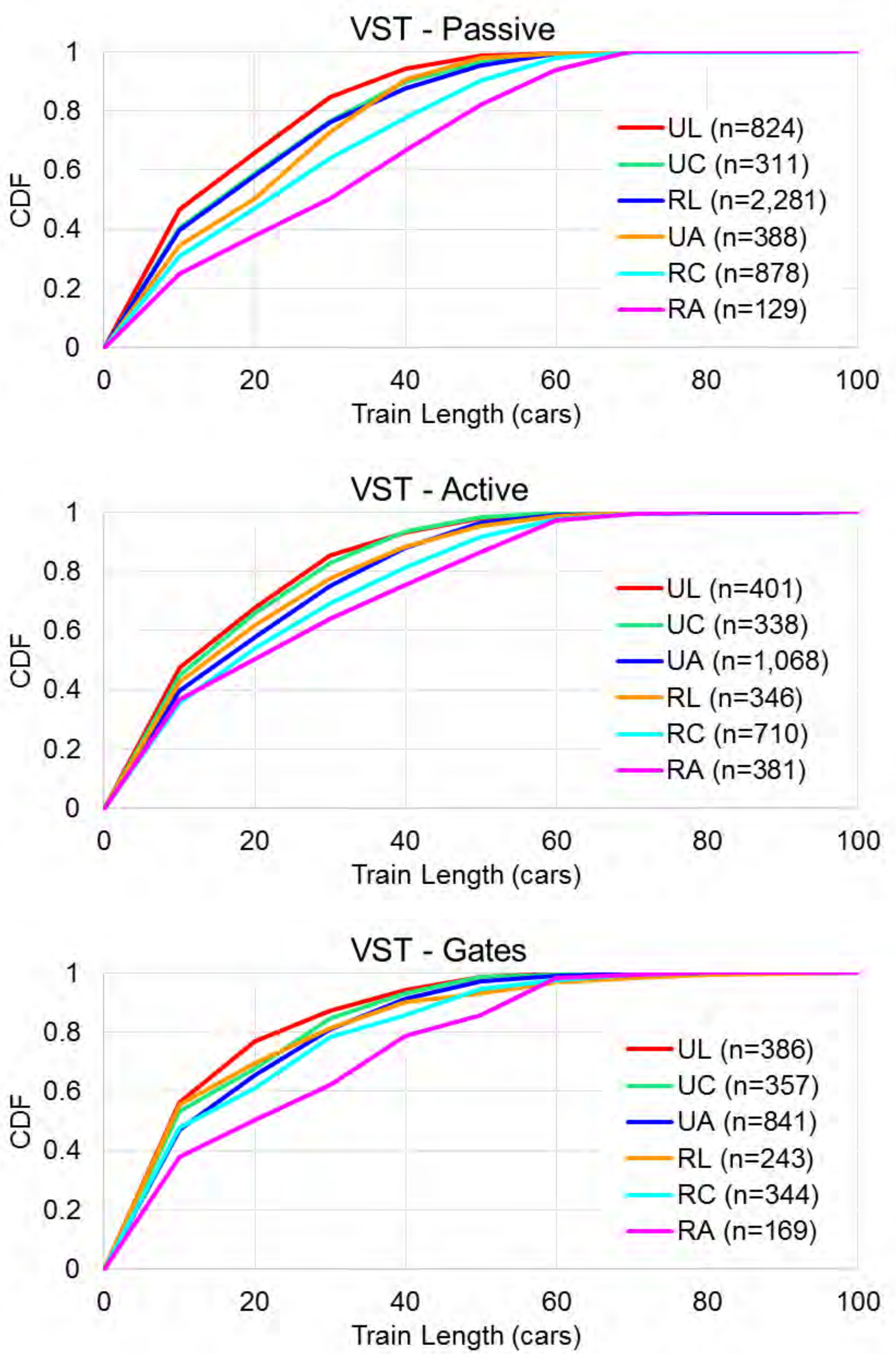


Figure D.1: Speed distributions for VST incidents, for (top) passive, (middle) active, and (bottom) gated crossings.

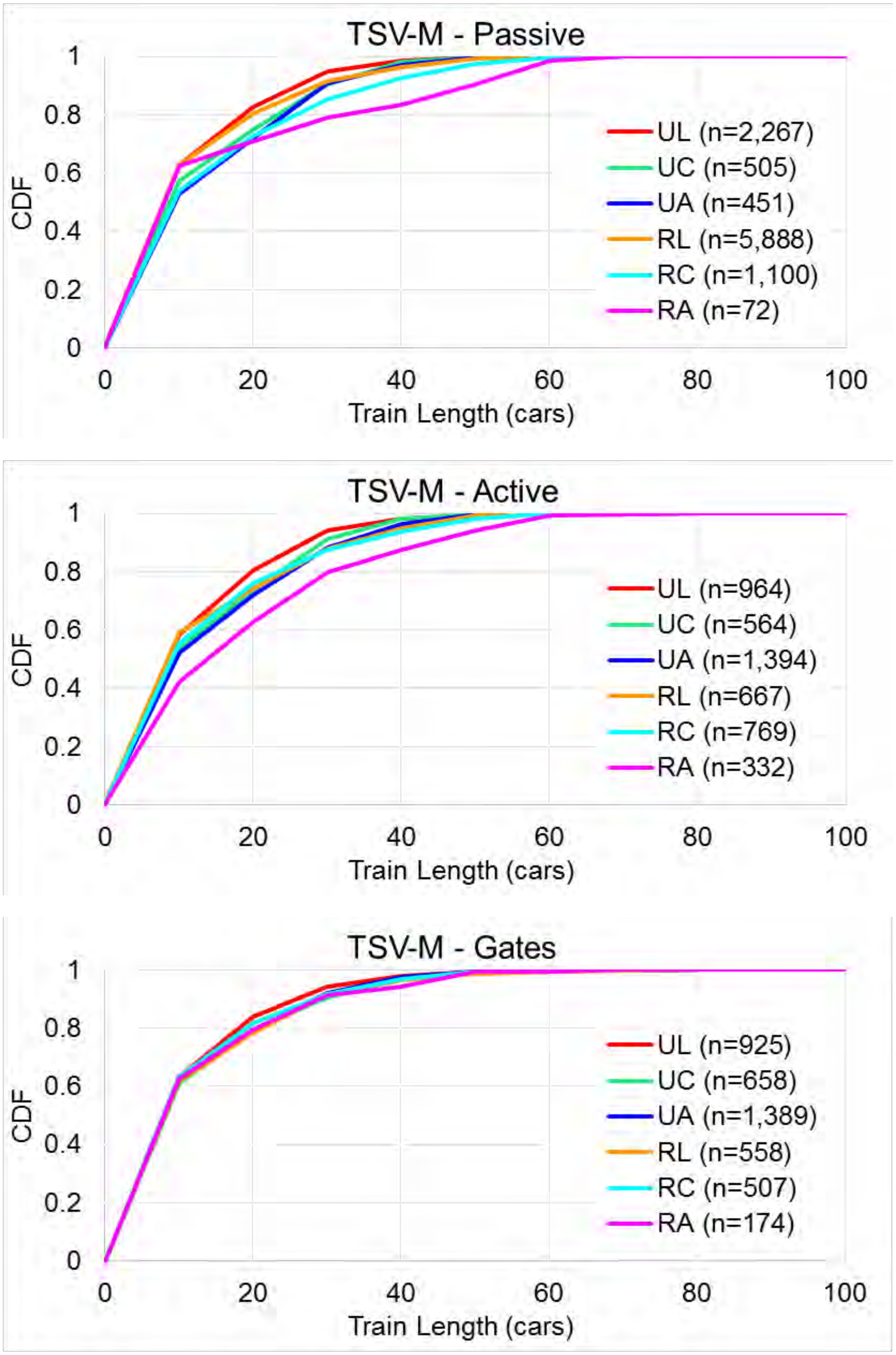


Figure D.2: Speed distributions for TSV-M incidents, for (top) passive, (middle) active, and (bottom) gated crossings.

APPENDIX E: HIGHWAY CLASSIFICATION CODES

Table E.1: Highway Classification Codes

Highway Class Group	GCI "HWYCLASS" Designation	GCI Description
Rural Arterial	2	Rural Principal Arterial
	6	Rural Minor Arterial
Rural Collector	7	Rural Major Collector
	8	Rural Minor Collector
Rural Local	9	Rural Local Road
Urban Arterial	14	Urban Other Principal Arterial
	16	Urban Minor Arterial
Urban Collector	17	Urban Collector
Urban Local	19	Urban Local Road

APPENDIX F: LIKELIHOOD OF TRUCK INVOLVEMENT

Figure F.1: Likelihood of Truck Involvement for Freight Consists

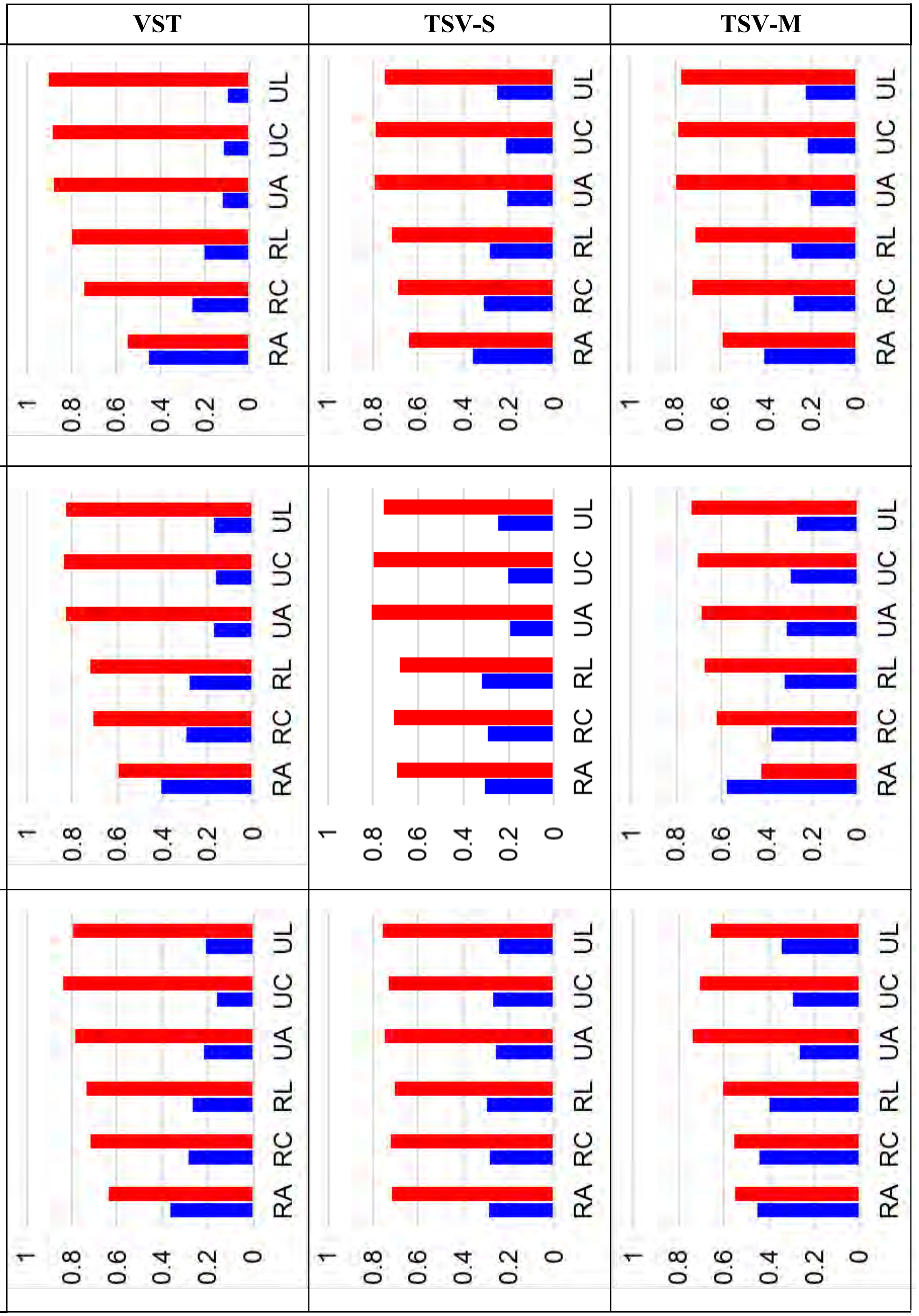
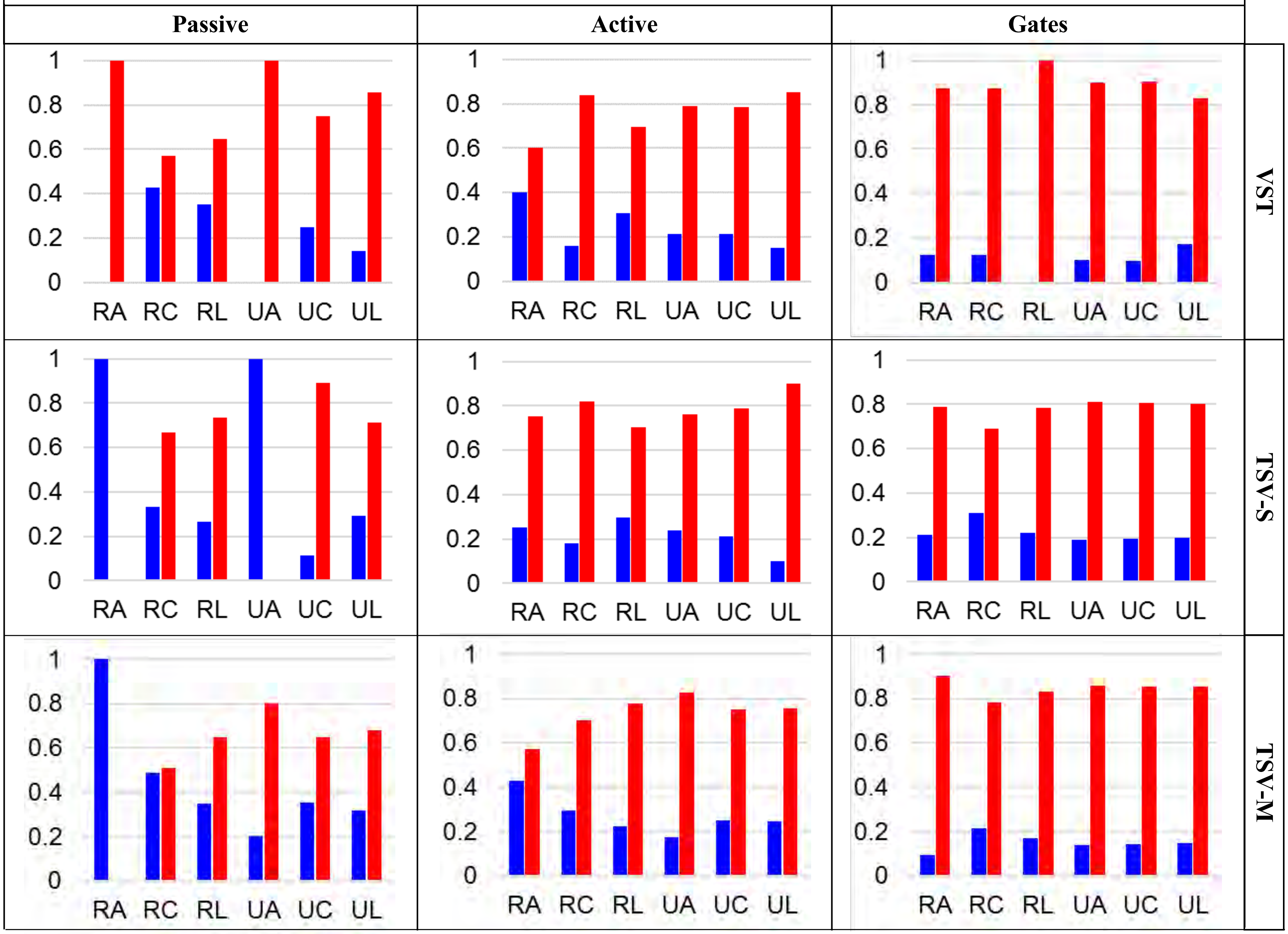


Figure F.2: Likelihood of Truck Involvement for Passenger Consists



APPENDIX G: DATA TABLES

Table G.1: Cumulative distribution of train speeds by incident type, track class and equipment type

TSV-M

Train Speed (mph)	Freight						Passenger					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.920	0.257	0.081	0.031	0.023	0.077	0.900	0.289	0.053	0.019	0.012	0.050
20	0.969	0.633	0.197	0.078	0.038	0.077	1.000	0.691	0.149	0.057	0.012	0.100
30	0.985	0.982	0.536	0.193	0.107	0.154	1.000	0.966	0.279	0.107	0.071	0.100
40	0.995	0.999	0.953	0.419	0.250	0.308	1.000	0.987	0.523	0.166	0.200	0.100
50	1.000	0.999	0.998	0.853	0.528	0.615	1.000	0.993	0.748	0.236	0.329	0.350
60	1.000	1.000	1.000	0.997	0.768	0.846	1.000	0.993	0.985	0.408	0.482	0.650
70	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	0.992	0.627	0.624	0.700
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.918	0.950
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.950
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0

TSV-S

Train Speed (mph)	Freight						Passenger					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.848	0.248	0.096	0.061	0.054	0.111	0.909	0.274	0.052	0.049	0.029	0.000
20	0.960	0.651	0.246	0.134	0.117	0.111	1.000	0.500	0.129	0.095	0.087	0.091
30	0.987	0.974	0.563	0.282	0.238	0.444	1.000	0.903	0.248	0.163	0.175	0.182
40	0.989	1.000	0.947	0.541	0.440	0.667	1.000	0.968	0.514	0.249	0.291	0.364
50	0.996	1.000	0.995	0.895	0.689	1.000	1.000	1.000	0.748	0.356	0.408	0.545
60	1.000	1.000	1.000	0.997	0.859	1.000	1.000	1.000	0.983	0.518	0.544	0.636
70	1.000	1.000	1.000	1.000	0.993	1.000	1.000	1.000	0.990	0.729	0.738	0.909
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.918	0.950
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.950
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table G.2: Large-to-small vehicle ratio by incident type, equipment type, warning device type and highway classification

Highway	Freight -- VST Incidents						Passenger -- VST Incidents					
	Gates		Other Active		Passive		Gates		Other Active		Passive	
Classification	LV	SV	LV	SV	LV	SV	LV	SV	LV	SV	LV	SV
RA	0.4533	0.5467	0.4045	0.5955	0.3652	0.6348	0.1250	0.8750	0.4000	0.6000	0.0000	1.0000
RC	0.2567	0.7433	0.2921	0.7079	0.2824	0.7176	0.1250	0.8750	0.1600	0.8400	0.4286	0.5714
RL	0.2010	0.7990	0.2800	0.7200	0.2666	0.7334	0.0000	1.0000	0.3043	0.6957	0.3529	0.6471
UA	0.1206	0.8794	0.1736	0.8264	0.2167	0.7833	0.1013	0.8987	0.2105	0.7895	0.0000	1.0000
UC	0.1143	0.8857	0.1630	0.8370	0.1597	0.8403	0.0952	0.9048	0.2143	0.7857	0.2500	0.7500
UL	0.0945	0.9055	0.1720	0.8280	0.2070	0.7930	0.1707	0.8293	0.1481	0.8519	0.1429	0.8571

Highway	Freight -- TSV-S Incidents						Passenger -- TSV-S Incidents					
	Gates		Other Active		Passive		Gates		Other Active		Passive	
Classification	LV	SV	LV	SV	LV	SV	LV	SV	LV	SV	LV	SV
RA	0.3594	0.6406	0.3048	0.6952	0.2857	0.7143	0.2105	0.7895	0.2500	0.7500	1.0000	0.0000
RC	0.3085	0.6915	0.2910	0.7090	0.2809	0.7191	0.3095	0.6905	0.1818	0.8182	0.3333	0.6667
RL	0.2817	0.7183	0.3181	0.6819	0.2963	0.7037	0.2182	0.7818	0.2963	0.7037	0.2667	0.7333
UA	0.2083	0.7917	0.1940	0.8060	0.2540	0.7460	0.1879	0.8121	0.2400	0.7600	1.0000	0.0000
UC	0.2102	0.7898	0.2023	0.7977	0.2690	0.7310	0.1938	0.8062	0.2105	0.7895	0.1111	0.8889
UL	0.2528	0.7472	0.2458	0.7542	0.2428	0.7572	0.2000	0.8000	0.1000	0.9000	0.2895	0.7105

Highway	Freight -- TSV-M Incidents						Passenger -- TSV-M Incidents					
	Gates		Other Active		Passive		Gates		Other Active		Passive	
Classification	LV	SV	LV	SV	LV	SV	LV	SV	LV	SV	LV	SV
RA	0.4087	0.5913	0.5769	0.4231	0.4500	0.5500	0.0952	0.9048	0.4286	0.5714	1.0000	0.0000
RC	0.2778	0.7222	0.3798	0.6202	0.4440	0.5560	0.2162	0.7838	0.2963	0.7037	0.4889	0.5111
RL	0.2879	0.7121	0.3234	0.6766	0.3961	0.6039	0.1705	0.8295	0.2241	0.7759	0.3509	0.6491
UA	0.2021	0.7979	0.3124	0.6876	0.2622	0.7378	0.1408	0.8592	0.1724	0.8276	0.2000	0.8000
UC	0.2145	0.7855	0.2943	0.7057	0.2949	0.7051	0.1453	0.8547	0.2500	0.7500	0.3529	0.6471
UL	0.2250	0.7750	0.2661	0.7339	0.3424	0.6576	0.1466	0.8534	0.2456	0.7544	0.3196	0.6804

Table G.3: Distribution of position struck in train by equipment type and highway vehicle size

Struck Position in Train	Freight				Passenger	
	Count		Percent of Total			
	Large Vehicles	Small Vehicles	Large Vehicles	Small Vehicles	All Vehicles	Percent of Total
1	1408	4448	0.56433	0.55565	381	0.75595
2	235	594	0.09419	0.07420	64	0.12698
3	96	195	0.03848	0.02436	27	0.05357
4	46	113	0.01844	0.01412	12	0.02381
5	36	126	0.01443	0.01574	1	0.00198
6	39	91	0.01563	0.01137	6	0.01190
7	29	70	0.01162	0.00874	4	0.00794
8	22	66	0.00882	0.00824	1	0.00198
9	15	49	0.00601	0.00612	1	0.00198
10	28	104	0.01122	0.01299	3	0.00595
11	19	42	0.00762	0.00525	0	0.00000
12	18	51	0.00721	0.00637	1	0.00198
13	15	46	0.00601	0.00575	0	0.00000
14	14	49	0.00561	0.00612	2	0.00397
15	13	69	0.00521	0.00862	0	0.00000
16	8	54	0.00321	0.00675	0	0.00000
17	9	42	0.00361	0.00525	0	0.00000
18	8	32	0.00321	0.00400	0	0.00000
19	12	32	0.00481	0.00400	0	0.00000
20	8	94	0.00321	0.01174	1	0.00198
21	12	29	0.00481	0.00362	0	0.00000
22	11	43	0.00441	0.00537	0	0.00000
23	12	34	0.00481	0.00425	0	0.00000
24	10	28	0.00401	0.00350	0	0.00000
25	15	45	0.00601	0.00562	0	0.00000
26	8	28	0.00321	0.00350	0	0.00000
27	9	35	0.00361	0.00437	0	0.00000
28	6	33	0.00240	0.00412	0	0.00000
29	5	27	0.00200	0.00337	0	0.00000
30	13	45	0.00521	0.00562	0	0.00000
31	6	32	0.00240	0.00400	0	0.00000
32	6	27	0.00240	0.00337	0	0.00000
33	8	21	0.00321	0.00262	0	0.00000
34	11	18	0.00441	0.00225	0	0.00000
35	11	28	0.00441	0.00350	0	0.00000
36	4	19	0.00160	0.00237	0	0.00000
37	4	16	0.00160	0.00200	0	0.00000
38	5	26	0.00200	0.00325	0	0.00000

Table G.3 (cont.)

Struck Position in Train	Freight				Passenger	
	Count		Percent of Total		All Vehicles	Percent of Total
	Large Vehicles	Small Vehicles	Large Vehicles	Small Vehicles		
39	6	18	0.00240	0.00225	0	0.00000
40	13	42	0.00521	0.00525	0	0.00000
41	4	25	0.00160	0.00312	0	0.00000
42	1	21	0.00040	0.00262	0	0.00000
43	2	16	0.00080	0.00200	0	0.00000
44	7	24	0.00281	0.00300	0	0.00000
45	7	22	0.00281	0.00275	0	0.00000
46	15	23	0.00601	0.00287	0	0.00000
47	6	20	0.00240	0.00250	0	0.00000
48	6	20	0.00240	0.00250	0	0.00000
49	6	15	0.00240	0.00187	0	0.00000
50	6	46	0.00240	0.00575	0	0.00000
51	5	18	0.00200	0.00225	0	0.00000
52	5	24	0.00200	0.00300	0	0.00000
53	4	9	0.00160	0.00112	0	0.00000
54	4	16	0.00160	0.00200	0	0.00000
55	4	19	0.00160	0.00237	0	0.00000
56	3	19	0.00120	0.00237	0	0.00000
57	5	17	0.00200	0.00212	0	0.00000
58	3	17	0.00120	0.00212	0	0.00000
59	2	13	0.00080	0.00162	0	0.00000
60	5	34	0.00200	0.00425	0	0.00000
61	7	18	0.00281	0.00225	0	0.00000
62	3	12	0.00120	0.00150	0	0.00000
63	5	16	0.00200	0.00200	0	0.00000
64	3	18	0.00120	0.00225	0	0.00000
65	5	22	0.00200	0.00275	0	0.00000
66	2	17	0.00080	0.00212	0	0.00000
67	2	21	0.00080	0.00262	0	0.00000
68	7	8	0.00281	0.00100	0	0.00000
69	1	8	0.00040	0.00100	0	0.00000
70	3	12	0.00120	0.00150	0	0.00000
71	4	16	0.00160	0.00200	0	0.00000
72	1	11	0.00040	0.00137	0	0.00000
73	4	11	0.00160	0.00137	0	0.00000
74	2	12	0.00080	0.00150	0	0.00000
75	4	22	0.00160	0.00275	0	0.00000
76	1	11	0.00040	0.00137	0	0.00000

Table G.3 (cont.)

Struck Position in Train	Freight				Passenger	
	Count		Percent of Total		All Vehicles	Percent of Total
	Large Vehicles	Small Vehicles	Large Vehicles	Small Vehicles		
77	2	8	0.00080	0.00100	0	0.00000
78	6	10	0.00240	0.00125	0	0.00000
79	1	14	0.00040	0.00175	0	0.00000
80	6	22	0.00240	0.00275	0	0.00000
81	3	14	0.00120	0.00175	0	0.00000
82	3	8	0.00120	0.00100	0	0.00000
83	1	9	0.00040	0.00112	0	0.00000
84	3	16	0.00120	0.00200	0	0.00000
85	4	17	0.00160	0.00212	0	0.00000
86	3	6	0.00120	0.00075	0	0.00000
87	5	4	0.00200	0.00050	0	0.00000
88	4	12	0.00160	0.00150	0	0.00000
89	2	12	0.00080	0.00150	0	0.00000
90	1	15	0.00040	0.00187	0	0.00000
91	4	6	0.00160	0.00075	0	0.00000
92	0	8	0.00000	0.00100	0	0.00000
93	2	11	0.00080	0.00137	0	0.00000
94	2	11	0.00080	0.00137	0	0.00000
95	2	8	0.00080	0.00100	0	0.00000
96	3	13	0.00120	0.00162	0	0.00000
97	0	9	0.00000	0.00112	0	0.00000
98	3	10	0.00120	0.00125	0	0.00000
99	0	10	0.00000	0.00125	0	0.00000
100+	28	137	0.01122	0.01711	0	0.00000

Table G.4: Distribution of number of locomotives in a train consist by equipment type and track class

Number of Locomotives	Freight					Passenger				
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 1	Class 2	Class 3	Class 4	Class 5
1	0.4167	0.3129	0.1217	0.0861	0.0447	0.7500	0.8333	0.7260	0.6085	0.5918
2	0.3750	0.4356	0.4646	0.4709	0.4124	0.1250	0.0556	0.1918	0.3298	0.2653
3	0.2083	0.1718	0.2721	0.2972	0.2887	0.1250	0.0000	0.0411	0.0468	0.0612
4	0.0000	0.0491	0.0996	0.1052	0.2027	0.0000	0.1111	0.0411	0.0128	0.0408
5	0.0000	0.0245	0.0310	0.0231	0.0309	0.0000	0.0000	0.0000	0.0021	0.0408
6	0.0000	0.0061	0.0066	0.0104	0.0069	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0022	0.0008	0.0069	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0022	0.0040	0.0034	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0024	0.0034	0.0000	0.0000	0.0000	0.0000	0.0000

Table G.5: Distribution of train lengths by track class and equipment type

Train Length	Track Class 1				Track Class 2			
	Count		Percent of Total		Count		Percent of Total	
(rail vehicles)	Passenger	Freight	Passenger	Freight	Passenger	Freight	Passenger	Freight
1 - 5	16	28	0.5926	0.0802	24	66	0.4706	0.0547
6 - 10	6	41	0.2222	0.1175	16	113	0.3137	0.0936
11 - 15	4	41	0.1481	0.1175	8	82	0.1569	0.0679
16 - 20	0	27	0.0000	0.0774	2	72	0.0392	0.0597
21 - 25	0	20	0.0000	0.0573	0	63	0.0000	0.0522
26 - 30	0	16	0.0000	0.0458	0	63	0.0000	0.0522
31 - 35	0	17	0.0000	0.0487	0	54	0.0000	0.0447
36 - 40	1	8	0.0370	0.0229	0	50	0.0000	0.0414
41 - 45	0	7	0.0000	0.0201	1	66	0.0196	0.0547
46 - 50	0	11	0.0000	0.0315	0	42	0.0000	0.0348
51 - 55	0	8	0.0000	0.0229	0	35	0.0000	0.0290
56 - 60	0	12	0.0000	0.0344	0	36	0.0000	0.0298
61 - 65	0	10	0.0000	0.0287	0	34	0.0000	0.0282
66 - 70	0	13	0.0000	0.0372	0	34	0.0000	0.0282
71 - 75	0	10	0.0000	0.0287	0	45	0.0000	0.0373
76 - 80	0	12	0.0000	0.0344	0	51	0.0000	0.0423
81 - 85	0	12	0.0000	0.0344	0	33	0.0000	0.0273
86 - 90	0	4	0.0000	0.0115	0	34	0.0000	0.0282
91 - 95	0	12	0.0000	0.0344	0	49	0.0000	0.0406
96 - 100	0	4	0.0000	0.0115	0	30	0.0000	0.0249
101 - 105	0	5	0.0000	0.0143	0	31	0.0000	0.0257
106 - 110	0	9	0.0000	0.0258	0	25	0.0000	0.0207
111 - 115	0	12	0.0000	0.0344	0	22	0.0000	0.0182
116 - 120	0	2	0.0000	0.0057	0	21	0.0000	0.0174
121 - 125	0	0	0.0000	0.0000	0	13	0.0000	0.0108
126 - 130	0	2	0.0000	0.0057	0	10	0.0000	0.0083
131 - 135	0	1	0.0000	0.0029	0	12	0.0000	0.0099
136 - 140	0	1	0.0000	0.0029	0	9	0.0000	0.0075
141 - 145	0	1	0.0000	0.0029	0	3	0.0000	0.0025
146 - 150	0	0	0.0000	0.0000	0	3	0.0000	0.0025
151 - 155	0	2	0.0000	0.0057	0	3	0.0000	0.0025
156 - 160	0	0	0.0000	0.0000	0	1	0.0000	0.0008
161 - 165	0	1	0.0000	0.0029	0	1	0.0000	0.0008
166 - 170	0	0	0.0000	0.0000	0	0	0.0000	0.0000
171 - 175	0	0	0.0000	0.0000	0	0	0.0000	0.0000
176 - 180	0	0	0.0000	0.0000	0	0	0.0000	0.0000
181 - 185	0	0	0.0000	0.0000	0	1	0.0000	0.0008
186 - 190	0	0	0.0000	0.0000	0	0	0.0000	0.0000
191 - 195	0	0	0.0000	0.0000	0	0	0.0000	0.0000
196 - 200	0	0	0.0000	0.0000	0	0	0.0000	0.0000
201 - 205	0	0	0.0000	0.0000	0	0	0.0000	0.0000

Table G.5 (cont.)

Length (rail vehicles)	Track Class 3				Track Class 4			
	Count		Percent of Total		Count		Percent of Total	
	Passenger	Freight	Passenger	Freight	Passenger	Freight	Passenger	Freight
1 - 5	76	66	0.2992	0.0257	278	82	0.3244	0.0163
6 - 10	132	103	0.5197	0.0401	293	146	0.3419	0.0290
11 - 15	33	111	0.1299	0.0432	177	141	0.2065	0.0280
16 - 20	9	84	0.0354	0.0327	61	165	0.0712	0.0328
21 - 25	2	115	0.0079	0.0447	20	171	0.0233	0.0340
26 - 30	0	92	0.0000	0.0358	8	228	0.0093	0.0453
31 - 35	1	98	0.0039	0.0381	4	217	0.0047	0.0431
36 - 40	0	107	0.0000	0.0416	2	241	0.0023	0.0479
41 - 45	0	110	0.0000	0.0428	6	246	0.0070	0.0489
46 - 50	1	103	0.0039	0.0401	4	202	0.0047	0.0401
51 - 55	0	101	0.0000	0.0393	2	226	0.0023	0.0449
56 - 60	0	90	0.0000	0.0350	0	232	0.0000	0.0461
61 - 65	0	106	0.0000	0.0412	0	232	0.0000	0.0461
66 - 70	0	109	0.0000	0.0424	0	238	0.0000	0.0473
71 - 75	0	97	0.0000	0.0377	0	203	0.0000	0.0403
76 - 80	0	118	0.0000	0.0459	0	211	0.0000	0.0419
81 - 85	0	98	0.0000	0.0381	0	193	0.0000	0.0383
86 - 90	0	109	0.0000	0.0424	0	166	0.0000	0.0330
91 - 95	0	133	0.0000	0.0518	0	182	0.0000	0.0361
96 - 100	0	98	0.0000	0.0381	1	179	0.0012	0.0356
101 - 105	0	105	0.0000	0.0409	1	179	0.0012	0.0356
106 - 110	0	90	0.0000	0.0350	0	162	0.0000	0.0322
111 - 115	0	76	0.0000	0.0296	0	150	0.0000	0.0298
116 - 120	0	67	0.0000	0.0261	0	180	0.0000	0.0357
121 - 125	0	50	0.0000	0.0195	0	118	0.0000	0.0234
126 - 130	0	48	0.0000	0.0187	0	114	0.0000	0.0226
131 - 135	0	22	0.0000	0.0086	0	74	0.0000	0.0147
136 - 140	0	19	0.0000	0.0074	0	67	0.0000	0.0133
141 - 145	0	13	0.0000	0.0051	0	19	0.0000	0.0038
146 - 150	0	10	0.0000	0.0039	0	20	0.0000	0.0040
151 - 155	0	7	0.0000	0.0027	0	31	0.0000	0.0062
156 - 160	0	5	0.0000	0.0019	0	3	0.0000	0.0006
161 - 165	0	1	0.0000	0.0004	0	9	0.0000	0.0018
166 - 170	0	5	0.0000	0.0019	0	0	0.0000	0.0000
171 - 175	0	3	0.0000	0.0012	0	4	0.0000	0.0008
176 - 180	0	1	0.0000	0.0004	0	1	0.0000	0.0002
181 - 185	0	0	0.0000	0.0000	0	0	0.0000	0.0000
186 - 190	0	0	0.0000	0.0000	0	0	0.0000	0.0000
191 - 195	0	0	0.0000	0.0000	0	2	0.0000	0.0004
196 - 200	0	0	0.0000	0.0000	0	1	0.0000	0.0002
201 - 205	0	0	0.0000	0.0000	0	0	0.0000	0.0000

Table G.5 (cont.)

Length (rail vehicles)	Track Class 5			
	Count		Percent of Total	
	Passenger	Freight	Passenger	Freight
1 - 5	12	10	0.1395	0.0122
6 - 10	54	10	0.6279	0.0122
11 - 15	16	15	0.1860	0.0183
16 - 20	1	17	0.0116	0.0207
21 - 25	0	22	0.0000	0.0268
26 - 30	2	30	0.0233	0.0366
31 - 35	0	25	0.0000	0.0305
36 - 40	0	45	0.0000	0.0549
41 - 45	0	37	0.0000	0.0451
46 - 50	0	36	0.0000	0.0439
51 - 55	0	40	0.0000	0.0488
56 - 60	0	41	0.0000	0.0500
61 - 65	0	41	0.0000	0.0500
66 - 70	1	59	0.0116	0.0720
71 - 75	0	70	0.0000	0.0854
76 - 80	0	56	0.0000	0.0683
81 - 85	0	34	0.0000	0.0415
86 - 90	0	19	0.0000	0.0232
91 - 95	0	20	0.0000	0.0244
96 - 100	0	23	0.0000	0.0280
101 - 105	0	34	0.0000	0.0415
106 - 110	0	30	0.0000	0.0366
111 - 115	0	32	0.0000	0.0390
116 - 120	0	30	0.0000	0.0366
121 - 125	0	9	0.0000	0.0110
126 - 130	0	14	0.0000	0.0171
131 - 135	0	7	0.0000	0.0085
136 - 140	0	9	0.0000	0.0110
141 - 145	0	3	0.0000	0.0037
146 - 150	0	0	0.0000	0.0000
151 - 155	0	1	0.0000	0.0012
156 - 160	0	1	0.0000	0.0012
161 - 165	0	0	0.0000	0.0000
166 - 170	0	0	0.0000	0.0000
171 - 175	0	0	0.0000	0.0000
176 - 180	0	0	0.0000	0.0000
181 - 185	0	0	0.0000	0.0000
186 - 190	0	0	0.0000	0.0000
191 - 195	0	0	0.0000	0.0000
196 - 200	0	0	0.0000	0.0000
201 - 205	0	0	0.0000	0.0000

APPENDIX H: SUPPLEMENTARY FILES

The supplementary file “Form_54_REA_DataStructure” is a .pdf file containing the datafile structure and field input specifications for the Federal Railroad Administration (FRA) Rail Equipment Accident/Incident database.

The supplementary file “Form_57_HRA_DataStructure” is a .pdf file containing the datafile structure and field input specifications for the FRA Highway-Rail Grade Crossing Accident/Incident database.

The supplementary file “GX_Inventory_FileStructure” is a .pdf file containing the data file structure and field input specifications for the U.S. DOT Crossing Inventory Form.

These files were downloaded from the FRA Office of Safety website in 2015. They are included with this dissertation because the fields and associated file structures in the databases are periodically updated.

(Note: Copies of these documents can be found at the end of this dissertation, after all appendices.)

APPENDIX I: STATISTICAL APPENDIX

I.1. CHAPTER 5 – FREIGHT MODEL

I.1.1. Unified Model (Section 5.4.1)

Input

```
title 'Stepwise regression on freight set c2'; *best;
proc logistic data=work.freight_c2 outest=betas covout;
class lgveh trnstk ;
model derail(event="1")= trnstk lgveh vehspd2 trnspd2 / selection=stepwise
waldcl clparm=wald ctable lackfit;
score data=work.freight_all fitstat;
run ;;
```

Output

Stepwise regression on freight set c2

The LOGISTIC Procedure

Model Information

Data Set	WORK.FREIGHT_C2	
Response Variable	DERAIL	DERAIL
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	996
Number of Observations Used	996

Response Profile

Ordered Value	DERAIL	Total Frequency
1	0	664
2	1	332

Probability modeled is DERAIL=1.

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables
LGVEH	N	1
	Y	-1
TRNSTK	N	1
	Y	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L = 1267.936

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
145.9513	4	<.0001

Step 1. Effect VEHSPD2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1269.936	1166.291
SC	1274.840	1176.098
-2 Log L	1267.936	1162.291

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	105.6453	1	<.0001
Score	109.3188	1	<.0001
Wald	90.9321	1	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
39.4155	3	<.0001

Note: No effects for the model in Step 1 are removed.

Step 2. Effect LGVEH entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1269.936	1151.122
SC	1274.840	1165.833
-2 Log L	1267.936	1145.122

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	122.8141	2	<.0001
Score	123.5302	2	<.0001
Wald	101.9251	2	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
23.6777	2	<.0001

Note: No effects for the model in Step 2 are removed.

Step 3. Effect TRNSTK entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1269.936	1137.059
SC	1274.840	1156.674
-2 Log L	1267.936	1129.059

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	138.8768	3	<.0001
Score	139.4060	3	<.0001
Wald	113.2547	3	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
7.1074	1	0.0077

Note: No effects for the model in Step 3 are removed.

Step 4. Effect TRNSPD2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1269.936	1131.981
SC	1274.840	1156.500
-2 Log L	1267.936	1121.981

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	145.9548	4	<.0001
Score	145.9513	4	<.0001
Wald	117.7824	4	<.0001

Note: No effects for the model in Step 4 are removed.

Note: All effects have been entered into the model.

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	VEHSPD2		1	1	109.3188		<.0001	VEHSPD2
2	LGVEH		1	2	16.1397		<.0001	LGVEH
3	TRNSTK		1	3	16.7521		<.0001	TRNSTK
4	TRNSPD2		1	4	7.1074		0.0077	TRNSPD2

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
TRNSTK	1	13.0373	0.0003
LGVEH	1	16.2197	<.0001
VEHSPD2	1	35.5985	<.0001
TRNSPD2	1	7.0565	0.0079

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.6001	0.2470	5.9027	0.0151
TRNSTK	N	1	0.3822	0.1059	13.0373	0.0003
LGVEH	N	1	-0.4106	0.1019	16.2197	<.0001
VEHSPD2		1	0.0316	0.00530	35.5985	<.0001
TRNSPD2		1	-0.0141	0.00533	7.0565	0.0079

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
TRNSTK N vs Y	2.148	1.418	3.252
LGVEH N vs Y	0.440	0.295	0.656
VEHSPD2	1.032	1.021	1.043
TRNSPD2	0.986	0.976	0.996

Association of Predicted Probabilities and Observed Responses

Percent Concordant	69.8	Somers' D	0.401
Percent Discordant	29.7	Gamma	0.403
Percent Tied	0.6	Tau-a	0.178
Pairs	220448	c	0.700

Parameter Estimates and Wald Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Intercept		-0.6001	-1.0842	-0.1160
TRNSTK	N	0.3822	0.1747	0.5897
LGVEH	N	-0.4106	-0.6104	-0.2108
VEHSPD2		0.0316	0.0212	0.0420
TRNSPD2		-0.0141	-0.0246	-0.00371

Partition for the Hosmer and Lemeshow Test

Group	Total	DERAIL = 1		DERAIL = 0	
		Observed	Expected	Observed	Expected
1	100	12	12.46	88	87.54
2	101	20	19.25	81	81.75
3	100	29	22.78	71	77.22
4	100	27	24.42	73	75.58
5	101	24	26.61	77	74.39

6	98	21	28.16	77	69.84
7	100	31	31.77	69	68.23
8	100	36	38.91	64	61.09
9	100	55	54.11	45	45.89
10	96	77	73.53	19	22.47

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
6.6297	8	0.5771

Classification Table

Prob	Correct		Incorrect		Percentages				
Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.080	332	0	664	0	33.3	100.0	0.0	66.7	.
0.100	330	6	658	2	33.7	99.4	0.9	66.6	25.0
0.120	325	33	631	7	35.9	97.9	5.0	66.0	17.5
0.140	320	71	593	12	39.3	96.4	10.7	65.0	14.5
0.160	316	99	565	16	41.7	95.2	14.9	64.1	13.9
0.180	314	119	545	18	43.5	94.6	17.9	63.4	13.1
0.200	307	137	527	25	44.6	92.5	20.6	63.2	15.4
0.220	295	170	494	37	46.7	88.9	25.6	62.6	17.9
0.240	265	250	414	67	51.7	79.8	37.7	61.0	21.1
0.260	232	342	322	100	57.6	69.9	51.5	58.1	22.6
0.280	217	405	259	115	62.4	65.4	61.0	54.4	22.1
0.300	198	461	203	134	66.2	59.6	69.4	50.6	22.5
0.320	175	499	165	157	67.7	52.7	75.2	48.5	23.9
0.340	168	531	133	164	70.2	50.6	80.0	44.2	23.6
0.360	158	549	115	174	71.0	47.6	82.7	42.1	24.1
0.380	155	565	99	177	72.3	46.7	85.1	39.0	23.9
0.400	145	575	89	187	72.3	43.7	86.6	38.0	24.5
0.420	139	586	78	193	72.8	41.9	88.3	35.9	24.8
0.440	133	597	67	199	73.3	40.1	89.9	33.5	25.0
0.460	125	606	58	207	73.4	37.7	91.3	31.7	25.5
0.480	119	610	54	213	73.2	35.8	91.9	31.2	25.9
0.500	114	613	51	218	73.0	34.3	92.3	30.9	26.2
0.520	105	615	49	227	72.3	31.6	92.6	31.8	27.0
0.540	103	619	45	229	72.5	31.0	93.2	30.4	27.0
0.560	99	623	41	233	72.5	29.8	93.8	29.3	27.2
0.580	92	627	37	240	72.2	27.7	94.4	28.7	27.7
0.600	87	631	33	245	72.1	26.2	95.0	27.5	28.0
0.620	82	638	26	250	72.3	24.7	96.1	24.1	28.2
0.640	77	642	22	255	72.2	23.2	96.7	22.2	28.4
0.660	72	645	19	260	72.0	21.7	97.1	20.9	28.7
0.680	63	646	18	269	71.2	19.0	97.3	22.2	29.4
0.700	57	651	13	275	71.1	17.2	98.0	18.6	29.7
0.720	52	651	13	280	70.6	15.7	98.0	20.0	30.1
0.740	46	651	13	286	70.0	13.9	98.0	22.0	30.5
0.760	42	653	11	290	69.8	12.7	98.3	20.8	30.8
0.780	34	655	9	298	69.2	10.2	98.6	20.9	31.3
0.800	27	657	7	305	68.7	8.1	98.9	20.6	31.7
0.820	17	660	4	315	68.0	5.1	99.4	19.0	32.3
0.840	11	663	1	321	67.7	3.3	99.8	8.3	32.6
0.860	10	664	0	322	67.7	3.0	100.0	0.0	32.7
0.880	5	664	0	327	67.2	1.5	100.0	0.0	33.0
0.900	2	664	0	330	66.9	0.6	100.0	0.0	33.2
0.920	1	664	0	331	66.8	0.3	100.0	0.0	33.3
0.940	0	664	0	332	66.7	0.0	100.0	.	33.3

Fit Statistics for SCORE Data												
Data Set	**Total Frequency**	**Log Likelihood**	**Error Rate**	**AIC**	**AICC**	**BIC**	**SC**	**R-Square**	**Max-Rescaled R-Square**	**AUC**	**Brier Score**
WORK.FREIGHT_ALL	43696	-15201.2	0.0972	30412.36	30412.36	30455.78	30455.78	-0.84221	-10.3584	0.754898	0.097175

I.2. Split Model (Section 5.4.2)

I.2.1.1. Input

```
title 'Stepwise regression on freight set c1n'; *best;
proc logistic data=work.freight_c1n outest=betas covout;
class lgveh  ;
model derail(event="1")= lgveh vehspd2 vehspd2*vehspd2 trnspd2
trnspd2*trnspd2 lgveh*vehspd2 lgveh*trnspd2 / selection=stepwise waldcl
clparm=wald ctable details lackfit outroc=roc1;
score data=work.freight_an fitstat;
run ;;

title 'Stepwise regression on freight set c2y' ; *best;
proc logistic data=work.freight_c2y outest=betas covout;
class lgveh  ;
model derail(event="1")= lgveh vehspd2 vehspd2*vehspd2 trnspd2
trnspd2*trnspd2 lgveh*vehspd2 lgveh*trnspd2/ selection=stepwise waldcl
clparm=wald ctable details lackfit outroc=roc1;
score data=work.freight_ay fitstat;
run ;;
```

I.2.1.2. Output

Stepwise regression on freight set c1n

The LOGISTIC Procedure

Model Information		
Data Set	WORK.FREIGHT_C1N	
Response Variable	DERAIL	DERAIL
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	330
Number of Observations Used	330

Response Profile

Ordered Value	DERAIL	Total Frequency
1	0	220
2	1	110

Probability modeled is DERAIL=1.

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables
LGVEH	N	1
	Y	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L = 420.099

Note: Under full-rank parameterizations, Type 3 effect tests are replaced by joint tests. The joint test for an effect is a test that all the parameters associated with that effect are zero. Such joint tests might not be equivalent to Type 3 effect tests under GLM parameterization.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.6931	0.1168	35.2332	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
145.5167	7	<.0001

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
LGVEH	1	101.6950	<.0001
VEHSPD2	1	64.0986	<.0001
TRNSPD2	1	1.2611	0.2614

Step 1. Effect LGVEH entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.099	315.792
SC	425.898	323.390
-2 Log L	420.099	311.792

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	108.3075	1	<.0001
Score	101.6950	1	<.0001
Wald	81.0998	1	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	81.0998	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.8855	0.1527	33.6325	<.0001
LGVEH	N	1	-1.3750	0.1527	81.0998	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
LGVEH N vs Y	0.064	0.035	0.116

Association of Predicted Probabilities and Observed Responses

Percent Concordant	62.6	Somers' D	0.586
Percent Discordant	4.0	Gamma	0.880
Percent Tied	33.4	Tau-a	0.261
Pairs	24200	c	0.793

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
61.5805	6	<.0001

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	81.0998	<.0001

Note: No effects for the model in Step 1 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
VEHSPD2	1	49.6191	<.0001
TRNSPD2	1	0.3041	0.5814

Step 2. Effect VEHSPD2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.099	266.042
SC	425.898	277.439
-2 Log L	420.099	260.042

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	160.0573	2	<.0001
Score	138.4282	2	<.0001
Wald	85.7872	2	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	68.6283	<.0001
VEHSPD2	1	41.6378	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.5140	0.3180	62.4968	<.0001
LGVEH	N	1	-1.4430	0.1742	68.6283	<.0001
VEHSPD2		1	0.0578	0.00896	41.6378	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
LGVEH N vs Y	0.056	0.028	0.110
VEHSPD2	1.060	1.041	1.078

Association of Predicted Probabilities and Observed Responses

Percent Concordant	87.5	Somers' D	0.770
Percent Discordant	10.6	Gamma	0.785
Percent Tied	1.9	Tau-a	0.343
Pairs	24200	c	0.885

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
9.8623	5	0.0792

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	68.6283	<.0001
VEHSPD2	1	41.6378	<.0001

Note: No effects for the model in Step 2 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
VEHSPD2*VEHSPD2	1	8.2480	0.0041
TRNSPD2	1	0.6923	0.4054
VEHSPD2*LGVEH	1	0.9219	0.3370

Step 3. Effect VEHSPD2*VEHSPD2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.099	258.609
SC	425.898	273.806
-2 Log L	420.099	250.609

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	169.4899	3	<.0001
Score	143.7228	3	<.0001
Wald	77.8157	3	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	65.0136	<.0001
VEHSPD2	1	1.1976	0.2738
VEHSPD2*VEHSPD2	1	8.0992	0.0044

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.6815	0.4086	16.9338	<.0001
LGVEH	N	1	-1.5233	0.1889	65.0136	<.0001
VEHSPD2		1	-0.0359	0.0328	1.1976	0.2738
VEHSPD2*VEHSPD2		1	0.00159	0.000558	8.0992	0.0044

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
LGVEH N vs Y	0.048	0.023	0.100

Association of Predicted Probabilities and Observed Responses

Percent Concordant	87.1	Somers' D	0.762
Percent Discordant	11.0	Gamma	0.776
Percent Tied	1.9	Tau-a	0.340
Pairs	24200	c	0.881

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
1.9391	4	0.7470

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	65.0136	<.0001
VEHSPD2*VEHSPD2	1	8.0992	0.0044

Note: No effects for the model in Step 3 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
TRNSPD2	1	0.6109	0.4344
VEHSPD2*LGVEH	1	0.0525	0.8188

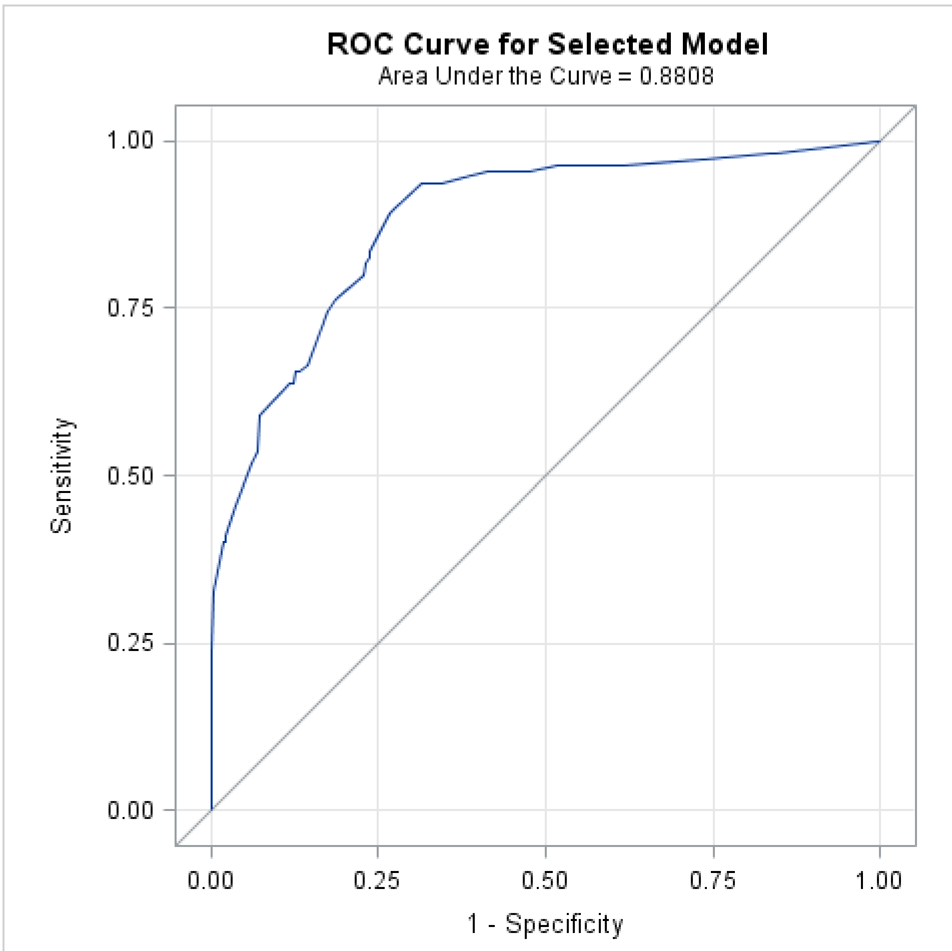
Note: No (additional) effects met the 0.05 significance level for entry into the model.

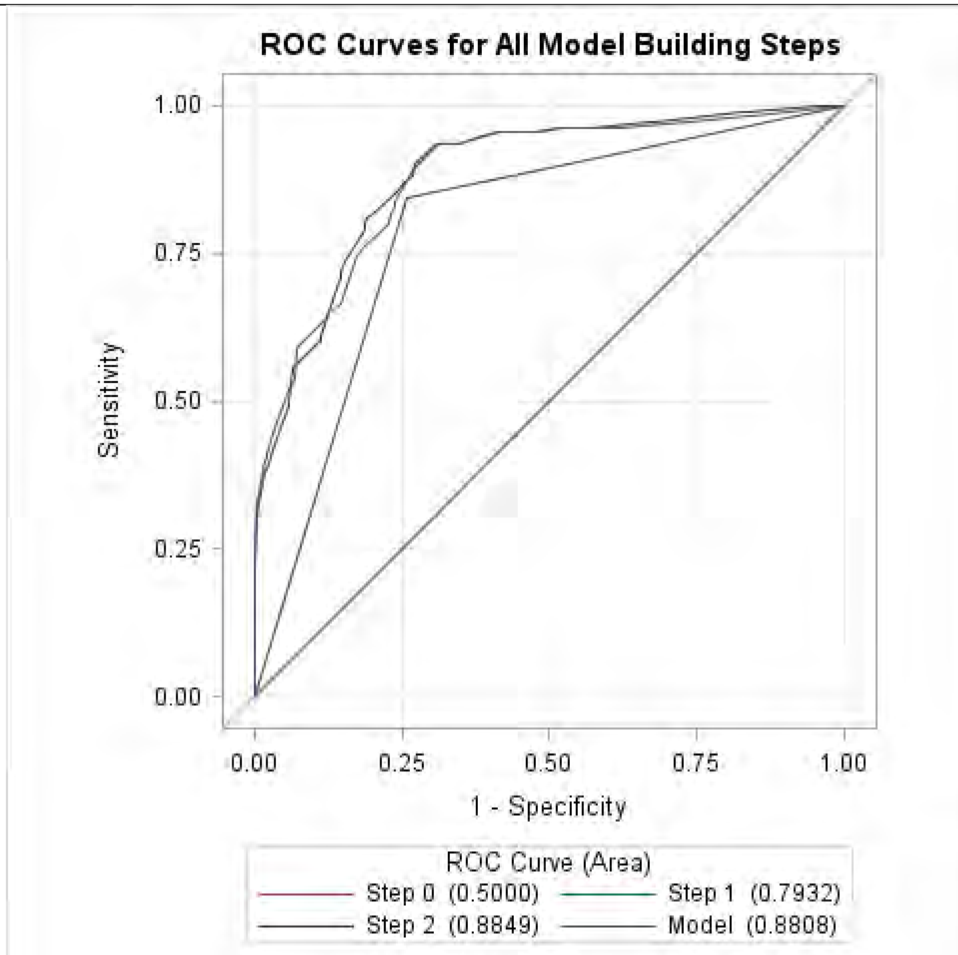
Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	LGVEH		1	1	101.6950		<.0001	LGVEH
2	VEHSPD2		1	2	49.6191		<.0001	VEHSPD2
3	VEHSPD2*VEHSPD2		1	3	8.2480		0.0041	

Parameter Estimates and Wald Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Intercept		-1.6815	-2.4824	-0.8806
LGVEH	N	-1.5233	-1.8936	-1.1530
VEHSPD2		-0.0359	-0.1002	0.0284
VEHSPD2*VEHSPD2		0.00159	0.000495	0.00268





Partition for the Hosmer and Lemeshow Test

Group	Total	DERAIL = 1		DERAIL = 0	
		Observed	Expected	Observed	Expected
1	36	2	1.16	34	34.84
2	25	1	0.82	24	24.18
3	30	1	1.03	29	28.97
4	29	1	1.12	28	27.88
5	31	2	2.31	29	28.69
6	35	11	10.95	24	24.05
7	39	19	16.82	20	22.18
8	35	16	17.51	19	17.49
9	33	21	23.72	12	9.28
10	37	36	34.56	1	2.44

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
3.5138	8	0.8981

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.020	110	0	220	0	33.3	100.0	0.0	66.7	.
0.040	105	106	114	5	63.9	95.5	48.2	52.1	4.5
0.060	105	129	91	5	70.9	95.5	58.6	46.4	3.7

0.080	104	137	83	6	73.0	94.5	62.3	44.4	4.2
0.100	103	137	83	7	72.7	93.6	62.3	44.6	4.9
0.120	103	144	76	7	74.8	93.6	65.5	42.5	4.6
0.140	103	144	76	7	74.8	93.6	65.5	42.5	4.6
0.160	103	144	76	7	74.8	93.6	65.5	42.5	4.6
0.180	103	150	70	7	76.7	93.6	68.2	40.5	4.5
0.200	103	150	70	7	76.7	93.6	68.2	40.5	4.5
0.220	103	151	69	7	77.0	93.6	68.6	40.1	4.4
0.240	103	151	69	7	77.0	93.6	68.6	40.1	4.4
0.260	99	151	69	11	75.8	90.0	68.6	41.1	6.8
0.280	99	159	61	11	78.2	90.0	72.3	38.1	6.5
0.300	99	159	61	11	78.2	90.0	72.3	38.1	6.5
0.320	99	159	61	11	78.2	90.0	72.3	38.1	6.5
0.340	99	159	61	11	78.2	90.0	72.3	38.1	6.5
0.360	99	159	61	11	78.2	90.0	72.3	38.1	6.5
0.380	99	159	61	11	78.2	90.0	72.3	38.1	6.5
0.400	98	159	61	12	77.9	89.1	72.3	38.4	7.0
0.420	84	166	54	26	75.8	76.4	75.5	39.1	13.5
0.440	70	179	41	40	75.5	63.6	81.4	36.9	18.3
0.460	70	192	28	40	79.4	63.6	87.3	28.6	17.2
0.480	65	193	27	45	78.2	59.1	87.7	29.3	18.9
0.500	65	204	16	45	81.5	59.1	92.7	19.8	18.1
0.520	65	204	16	45	81.5	59.1	92.7	19.8	18.1
0.540	65	204	16	45	81.5	59.1	92.7	19.8	18.1
0.560	59	205	15	51	80.0	53.6	93.2	20.3	19.9
0.580	57	205	15	53	79.4	51.8	93.2	20.8	20.5
0.600	57	205	15	53	79.4	51.8	93.2	20.8	20.5
0.620	57	205	15	53	79.4	51.8	93.2	20.8	20.5
0.640	50	212	8	60	79.4	45.5	96.4	13.8	22.1
0.660	50	212	8	60	79.4	45.5	96.4	13.8	22.1
0.680	50	212	8	60	79.4	45.5	96.4	13.8	22.1
0.700	50	212	8	60	79.4	45.5	96.4	13.8	22.1
0.720	44	212	8	66	77.6	40.0	96.4	15.4	23.7
0.740	44	215	5	66	78.5	40.0	97.7	10.2	23.5
0.760	44	215	5	66	78.5	40.0	97.7	10.2	23.5
0.780	44	216	4	66	78.8	40.0	98.2	8.3	23.4
0.800	44	216	4	66	78.8	40.0	98.2	8.3	23.4
0.820	36	219	1	74	77.3	32.7	99.5	2.7	25.3
0.840	36	219	1	74	77.3	32.7	99.5	2.7	25.3
0.860	36	219	1	74	77.3	32.7	99.5	2.7	25.3
0.880	35	219	1	75	77.0	31.8	99.5	2.8	25.5
0.900	25	220	0	85	74.2	22.7	100.0	0.0	27.9
0.920	24	220	0	86	73.9	21.8	100.0	0.0	28.1
0.940	12	220	0	98	70.3	10.9	100.0	0.0	30.8
0.960	12	220	0	98	70.3	10.9	100.0	0.0	30.8
0.980	9	220	0	101	69.4	8.2	100.0	0.0	31.5
1.000	0	220	0	110	66.7	0.0	100.0	.	33.3

Fit Statistics for SCORE Data

Data Set	Total Frequency	Log Likelihood	Error Rate	AIC	AICC	BIC	SC	R-Square	Max-Rescaled R-Square	AUC	Brier Score
WORK.FREIGHT_AN	8753	-2678.1	0.0993	5364.215	5364.219	5392.523	5392.523	-0.61115	-4.84017	0.876037	0.091264

Stepwise regression on freight set c2y

The LOGISTIC Procedure

Model Information

Data Set	WORK.FREIGHT_C2Y	
Response Variable	DERAIL	DERAIL
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	606

Number of Observations Used	606

Response Profile

Ordered Value	DERAIL	Total Frequency
1	0	404
2	1	202

Probability modeled is DERAIL=1.

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables
LGVEH	**N**	1
	Y	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L = 771.455

Note: Under full-rank parameterizations, Type 3 effect tests are replaced by joint tests. The joint test for an effect is a test that all the parameters associated with that effect are zero. Such joint tests might not be equivalent to Type 3 effect tests under GLM parameterization.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.6931	0.0862	64.7010	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
239.0867	7	<.0001

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
LGVEH	1	223.5137	<.0001
VEHSPD2	1	10.0964	0.0015
TRNSPD2	1	13.7307	0.0002

Step 1. Effect LGVEH entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	773.455	514.318
SC	777.862	523.132
-2 Log L	771.455	510.318

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	261.1371	1	<.0001
Score	223.5137	1	<.0001
Wald	121.6025	1	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	121.6025	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.4475	0.1709	71.7278	<.0001
LGVEH	N	1	-1.8847	0.1709	121.6025	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
LGVEH N vs Y	0.023	0.012	0.045

Association of Predicted Probabilities and Observed Responses

Percent Concordant	65.9	Somers' D	0.644
Percent Discordant	1.5	Gamma	0.955
Percent Tied	32.6	Tau-a	0.287
Pairs	81608	c	0.822

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
16.6537	6	0.0106

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	121.6025	<.0001

Note: No effects for the model in Step 1 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
VEHSPD2	1	2.5173	0.1126
TRNSPD2	1	4.7511	0.0293

Step 2. Effect TRNSPD2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	773.455	511.554
SC	777.862	524.775
-2 Log L	771.455	505.554

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	265.9011	2	<.0001
Score	226.4661	2	<.0001
Wald	123.8537	2	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	119.1551	<.0001
TRNSPD2	1	4.7008	0.0301

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.0330	0.3239	39.4008	<.0001
LGVEH	N	1	-1.8687	0.1712	119.1551	<.0001
TRNSPD2		1	0.0166	0.00766	4.7008	0.0301

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
LGVEH N vs Y	0.024	0.012	0.047
TRNSPD2	1.017	1.002	1.032

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.7	Somers' D	0.684
Percent Discordant	15.3	Gamma	0.691
Percent Tied	0.9	Tau-a	0.305
Pairs	81608	c	0.842

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
11.4650	5	0.0429

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
LGVEH	1	119.1551	<.0001
TRNSPD2	1	4.7008	0.0301

Note: No effects for the model in Step 2 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
VEHSPD2	1	3.6351	0.0566
TRNSPD2*TRNSPD2	1	1.6524	0.1986
TRNSPD2*LGVEH	1	1.1214	0.2896

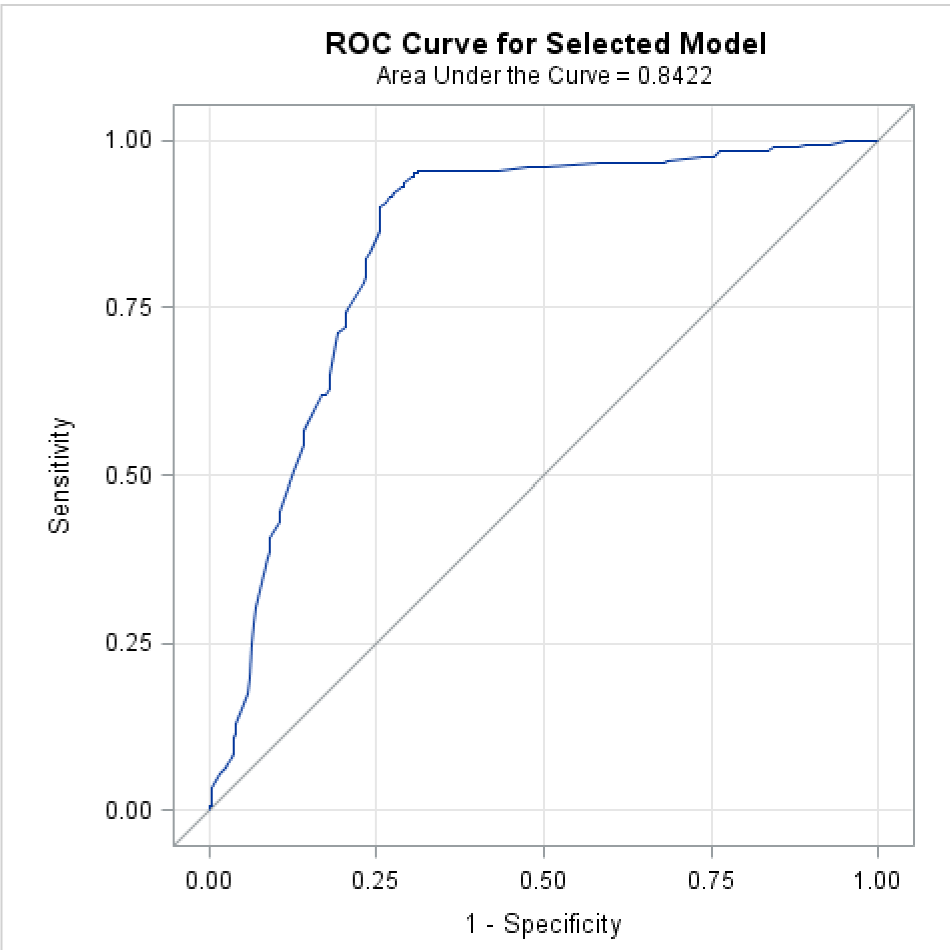
Note: No (additional) effects met the 0.05 significance level for entry into the model.

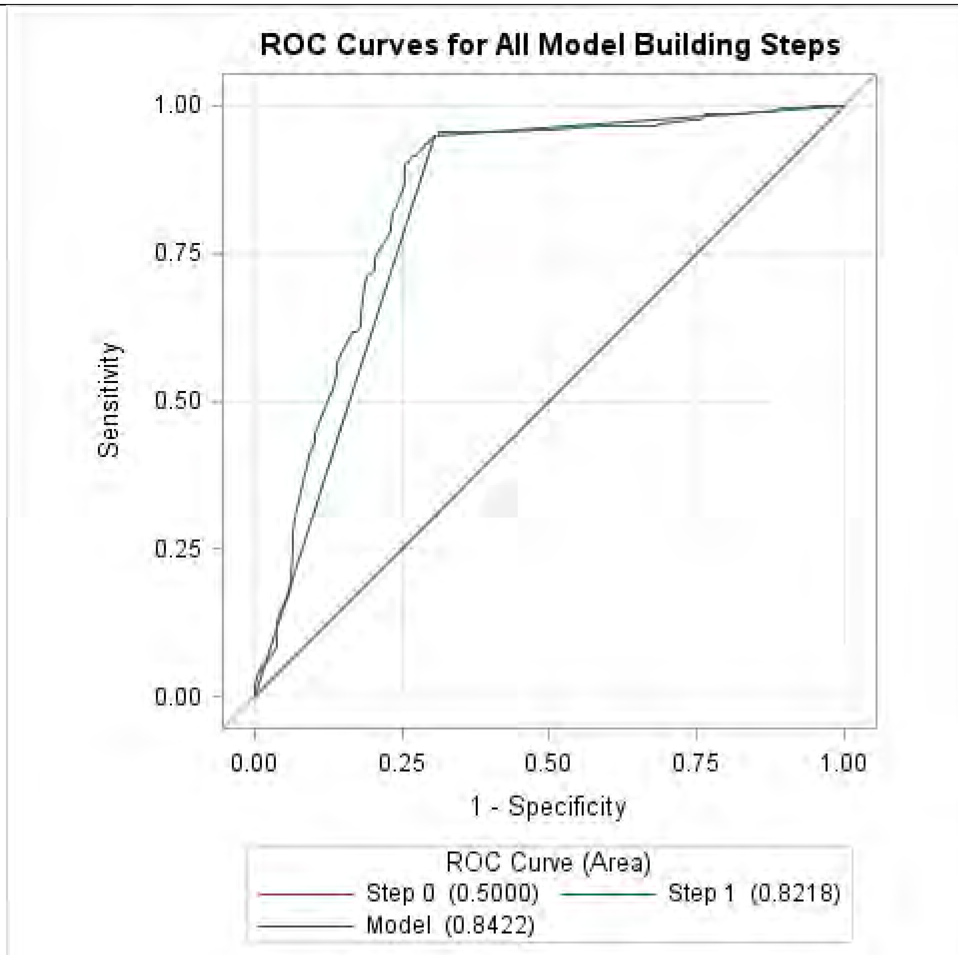
Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	**LGVEH**		1	1	223.5137		<.0001	LGVEH
2	**TRNSPD2**		1	2	4.7511		0.0293	TRNSPD2

Parameter Estimates and Wald Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Intercept		-2.0330	-2.6677	-1.3982
LGVEH	**N**	-1.8687	-2.2042	-1.5331
TRNSPD2		0.0166	0.00159	0.0316





Partition for the Hosmer and Lemeshow Test

Group	Total	DERAIL = 1		DERAIL = 0	
		Observed	Expected	Observed	Expected
1	65	2	1.58	63	63.42
2	68	4	2.07	64	65.93
3	76	2	2.77	74	73.23
4	62	1	2.59	61	59.41
5	57	18	20.25	39	36.75
6	61	34	34.13	27	26.87
7	64	39	38.45	25	25.55
8	64	41	40.45	23	23.55
9	89	61	59.71	28	29.29

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
3.7232	7	0.8110

Classification Table

Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.020	202	0	404	0	33.3	100.0	0.0	66.7	.
0.040	193	209	195	9	66.3	95.5	51.7	50.3	4.1
0.060	192	278	126	10	77.6	95.0	68.8	39.6	3.5
0.080	192	280	124	10	77.9	95.0	69.3	39.2	3.4

0.100	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.120	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.140	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.160	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.180	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.200	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.220	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.240	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.260	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.280	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.300	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.320	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.340	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.360	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.380	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.400	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.420	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.440	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.460	192	280	124	10	77.9	95.0	69.3	39.2	3.4
0.480	189	281	123	13	77.6	93.6	69.6	39.4	4.4
0.500	186	287	117	16	78.1	92.1	71.0	38.6	5.3
0.520	183	294	110	19	78.7	90.6	72.8	37.5	6.1
0.540	168	301	103	34	77.4	83.2	74.5	38.0	10.1
0.560	159	310	94	43	77.4	78.7	76.7	37.2	12.2
0.580	141	327	77	61	77.2	69.8	80.9	35.3	15.7
0.600	125	334	70	77	75.7	61.9	82.7	35.9	18.7
0.620	102	350	54	100	74.6	50.5	86.6	34.6	22.2
0.640	76	368	36	126	73.3	37.6	91.1	32.1	25.5
0.660	26	379	25	176	66.8	12.9	93.8	49.0	31.7
0.680	13	389	15	189	66.3	6.4	96.3	53.6	32.7
0.700	6	398	6	196	66.7	3.0	98.5	50.0	33.0
0.720	1	403	1	201	66.7	0.5	99.8	50.0	33.3
0.740	0	404	0	202	66.7	0.0	100.0	.	33.3

Fit Statistics for SCORE Data

Data Set	Total Frequency	Log Likelihood	Error Rate	AIC	AICC	BIC	SC	R-Square	Max-Rescaled R-Square	AUC	Brier Score
WORK.FREIGHT_AY	34903	-11383.4	0.3031	22772.86	22772.86	22798.24	22798.24	-0.78803	-11.4702	0.83599	0.118878

I.3. CHAPTER 6 – PASSENGER MODEL

I.3.1. VST

Input

```
title 'logistic analysis for small VST2';
proc logistic data=work.sm_S_VST2 outest=betas covout plots=(roc);
class lgveh frtpax wt_class;
model der(event="1")=  vehspd2*lgveh vehspd2 lgveh trnspd2*wt_class trnspd2
wt_class totleng nbrloco/ selection=stepwise waldcl clparm=wald ctable
lackfit expb outroc=roc1;
output out=pred predprobs=(individual crossvalidate);
```

```
score data=work.VST fitstat;
run ;;
```

Output

logistic analysis for small VST4 incl totleng stepwise -- good model

The LOGISTIC Procedure

Model Information

Data Set	WORK.SM_S_VST2	
Response Variable	der	der
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	292
Number of Observations Used	292

Response Profile

Ordered Value	der	Total Frequency
1	0	195
2	1	97

Probability modeled is der=1.

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables		
LGVEH	**N**	1		
	Y	-1		
wt_class	**FRT_CAR**	1	0	0
	FRT_LOCO	0	1	0
	PAX_CAR	0	0	1
	PAX_LOCO	-1	-1	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L = 371.260

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
137.7850	12	<.0001

Step 1. Effect LGVEH entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	373.260	299.220
SC	376.937	306.573
-2 Log L	371.260	295.220

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	76.0410	1	<.0001
Score	72.6854	1	<.0001
Wald	61.7881	1	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
86.6937	11	<.0001

Note: No effects for the model in Step 1 are removed.

Step 2. Effect vehspd2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	373.260	246.039
SC	376.937	257.069
-2 Log L	371.260	240.039

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	131.2218	2	<.0001
Score	112.6715	2	<.0001
Wald	70.5053	2	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
36.9456	10	<.0001

Note: No effects for the model in Step 2 are removed.

Step 3. Effect wt_class entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	373.260	230.713
SC	376.937	252.773
-2 Log L	371.260	218.713

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	152.5479	5	<.0001
Score	125.4690	5	<.0001
Wald	72.0964	5	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
14.2688	7	0.0466

Note: No effects for the model in Step 3 are removed.

Note: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	LGVEH		1	1	72.6854		<.0001	LGVEH
2	vehspd2		1	2	53.5530		<.0001	vehspd2
3	wt_class		3	3	23.7066		<.0001	wt_class

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
vehspd2	1	37.7556	<.0001
LGVEH	1	55.1555	<.0001
wt_class	3	19.9865	0.0002

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-2.0204	0.4179	23.3777	<.0001	0.133
vehspd2		1	0.0607	0.00988	37.7556	<.0001	1.063
LGVEH	N	1	-1.5458	0.2081	55.1555	<.0001	0.213
wt_class	FRT_CAR	1	0.0648	0.3589	0.0326	0.8567	1.067
wt_class	FRT_LOCO	1	-1.3087	0.3737	12.2651	0.0005	0.270
wt_class	PAX_CAR	1	1.8213	0.7159	6.4720	0.0110	6.180

Odds Ratio Estimates

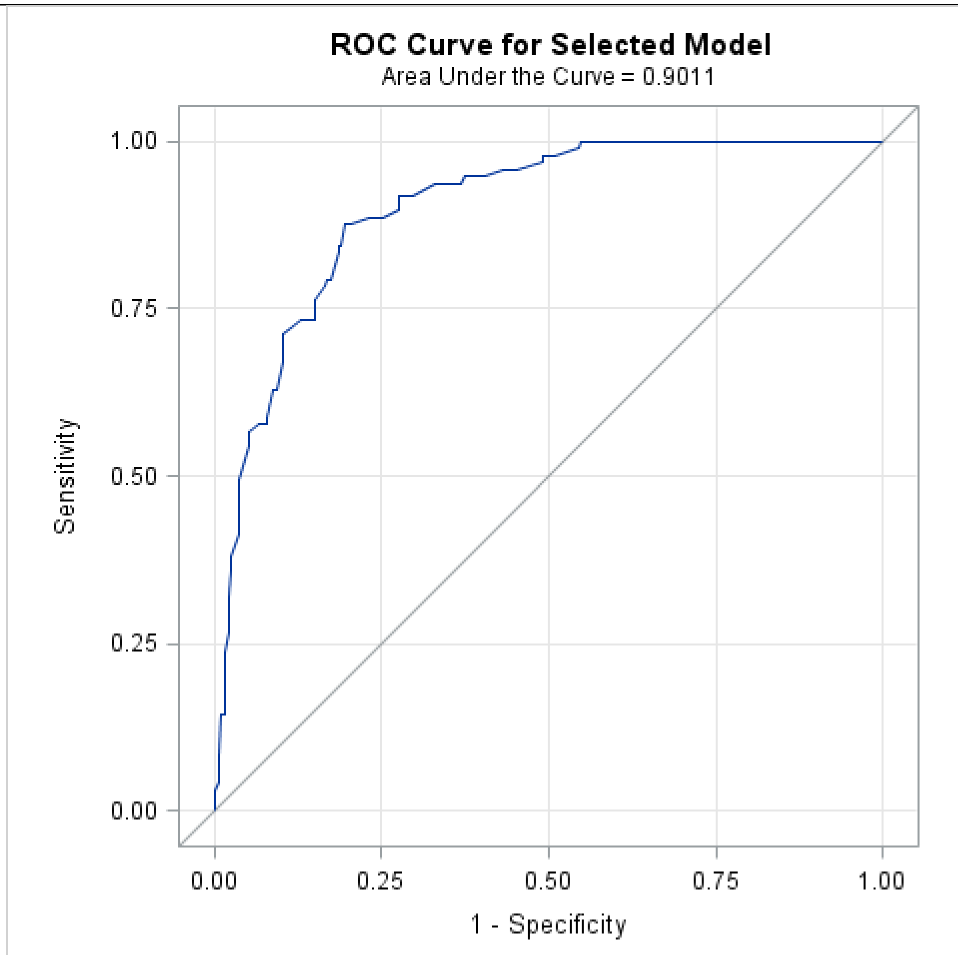
Effect	Point Estimate	95% Wald Confidence Limits	
vehspd2	1.063	1.042	1.083
LGVEH N vs Y	0.045	0.020	0.103
wt_class FRT_CAR vs PAX_LOCO	1.901	0.368	9.808
wt_class FRT_LOCO vs PAX_LOCO	0.481	0.093	2.486
wt_class PAX_CAR vs PAX_LOCO	11.010	1.041	116.493

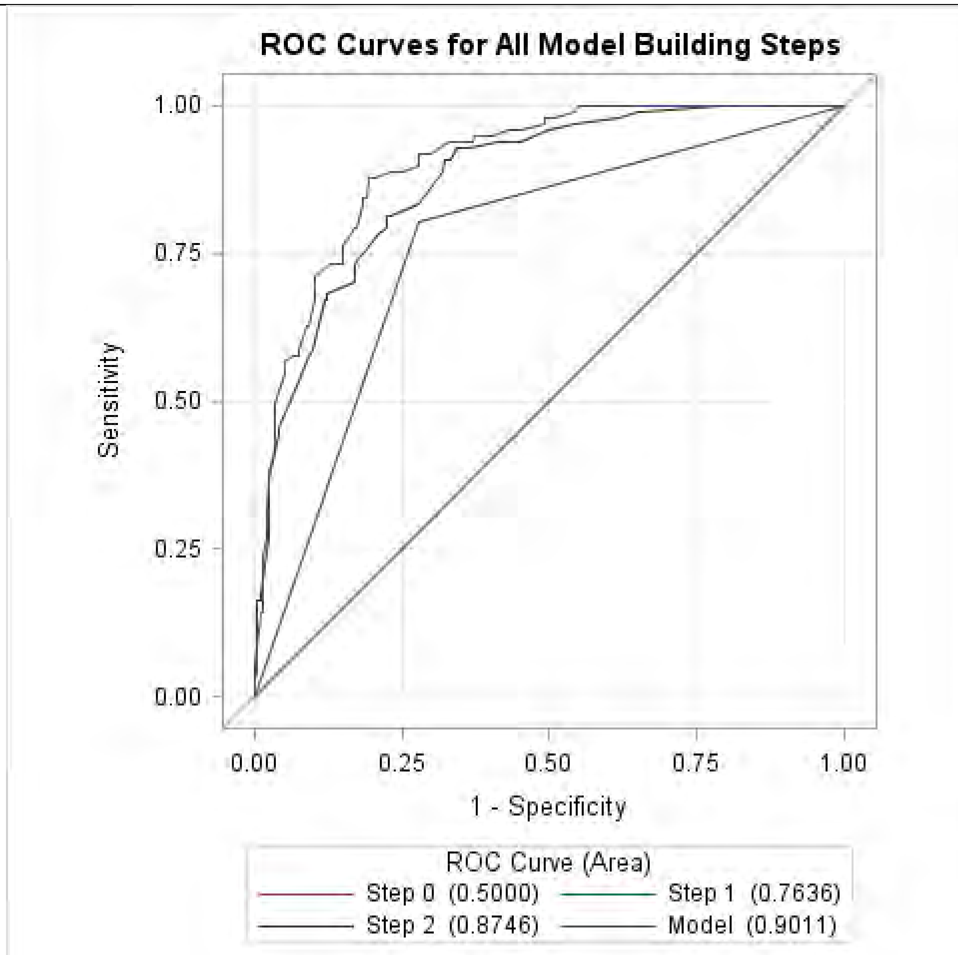
Association of Predicted Probabilities and Observed Responses

Percent Concordant	89.8	Somers' D	0.802
Percent Discordant	9.6	Gamma	0.808
Percent Tied	0.7	Tau-a	0.357
Pairs	18915	c	0.901

Parameter Estimates and Wald Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Intercept		-2.0204	-2.8394	-1.2014
vehspd2		0.0607	0.0413	0.0801
LGVEH	N	-1.5458	-1.9538	-1.1379
wt_class	FRT_CAR	0.0648	-0.6387	0.7683
wt_class	FRT_LOCO	-1.3087	-2.0411	-0.5763
wt_class	PAX_CAR	1.8213	0.4181	3.2245


ROC Curve for Selected Model
Area Under the Curve = 0.9011
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity



Partition for the Hosmer and Lemeshow Test

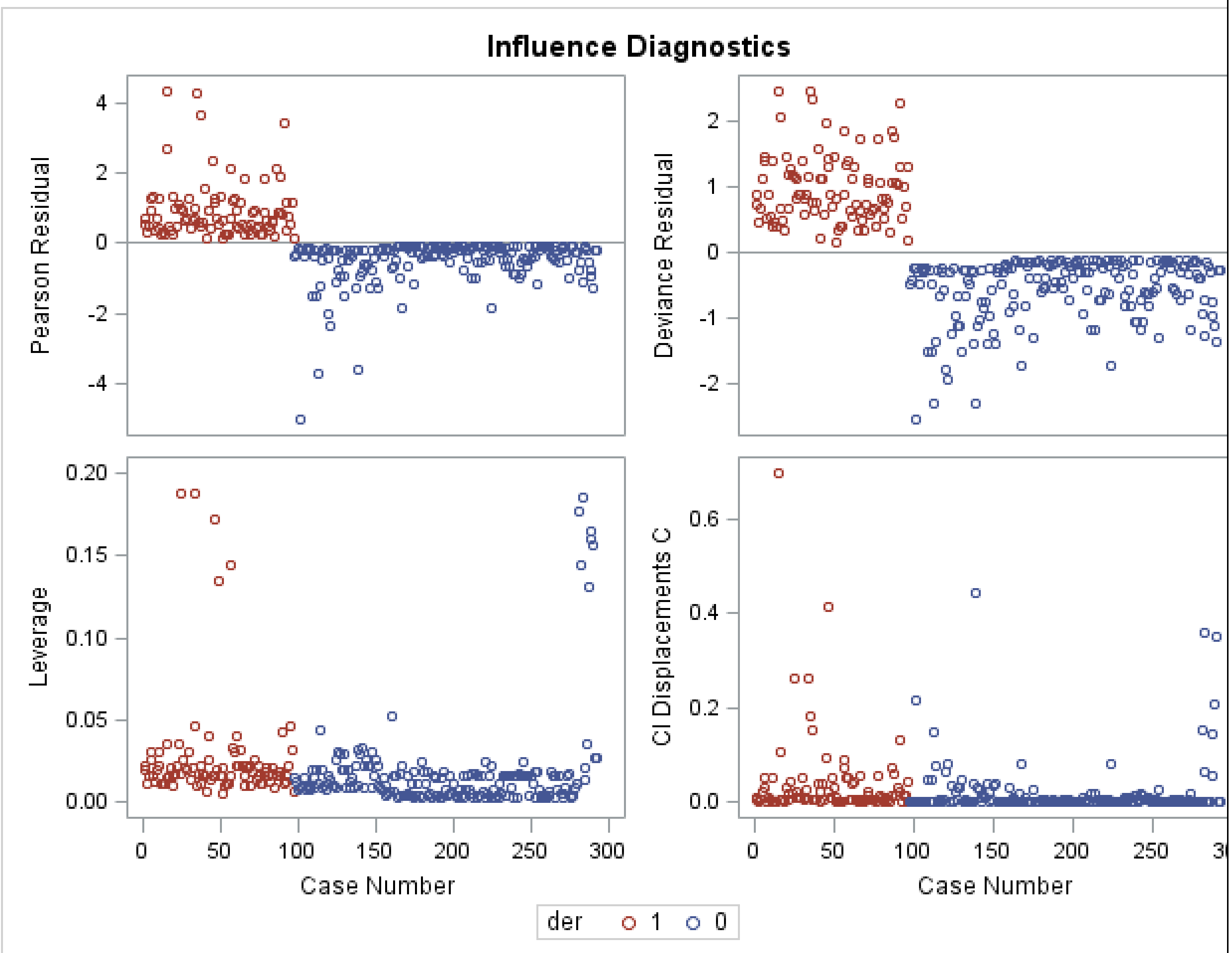
Group	Total	der = 1		der = 0	
		Observed	Expected	Observed	Expected
1	23	0	0.22	23	22.78
2	34	0	0.61	34	33.39
3	30	0	1.13	30	28.87
4	28	4	1.93	24	26.07
5	30	4	4.67	26	25.33
6	29	7	7.91	22	21.09
7	29	13	12.88	16	16.12
8	26	16	15.15	10	10.85
9	29	23	21.68	6	7.32
10	34	30	30.82	4	3.18

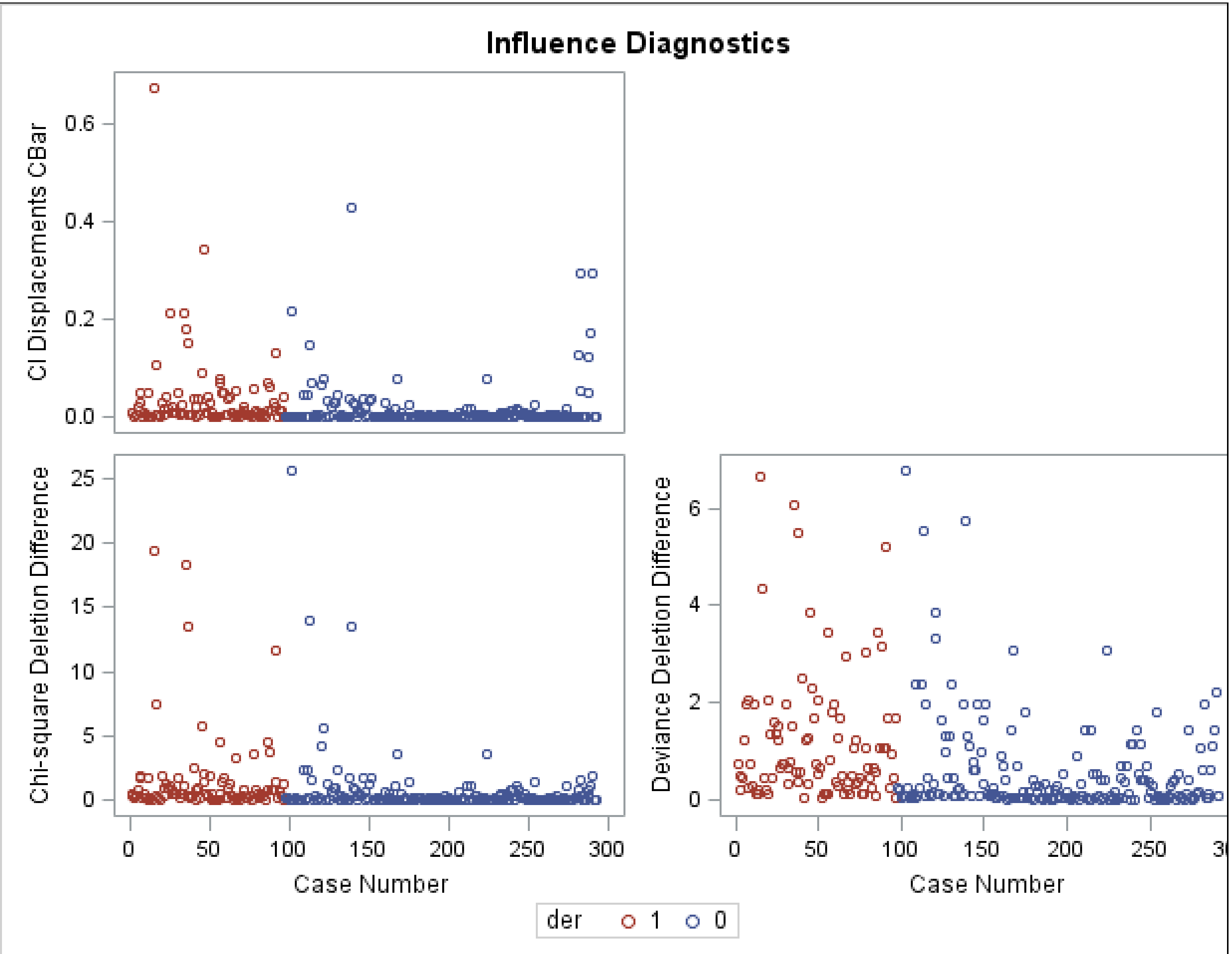
Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
5.3259	8	0.7222

Classification Table									
Prob	Correct		Incorrect		Percentages				
Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	97	0	195	0	33.2	100.0	0.0	66.8	.
0.020	97	47	148	0	49.3	100.0	24.1	60.4	0.0
0.040	96	80	115	1	60.3	99.0	41.0	54.5	1.2
0.060	95	96	99	2	65.4	97.9	49.2	51.0	2.0
0.080	93	99	96	4	65.8	95.9	50.8	50.8	3.9
0.100	93	109	86	4	69.2	95.9	55.9	48.0	3.5
0.120	92	111	84	5	69.5	94.8	56.9	47.7	4.3
0.140	91	121	74	6	72.6	93.8	62.1	44.8	4.7
0.160	91	125	70	6	74.0	93.8	64.1	43.5	4.6
0.180	89	128	67	8	74.3	91.8	65.6	42.9	5.9
0.200	89	137	58	8	77.4	91.8	70.3	39.5	5.5
0.220	87	141	54	10	78.1	89.7	72.3	38.3	6.6
0.240	86	146	49	11	79.5	88.7	74.9	36.3	7.0
0.260	85	146	49	12	79.1	87.6	74.9	36.6	7.6
0.280	85	148	47	12	79.8	87.6	75.9	35.6	7.5
0.300	84	155	40	13	81.8	86.6	79.5	32.3	7.7
0.320	84	155	40	13	81.8	86.6	79.5	32.3	7.7
0.340	84	156	39	13	82.2	86.6	80.0	31.7	7.7
0.360	81	156	39	16	81.2	83.5	80.0	32.5	9.3
0.380	77	158	37	20	80.5	79.4	81.0	32.5	11.2
0.400	76	160	35	21	80.8	78.4	82.1	31.5	11.6
0.420	74	160	35	23	80.1	76.3	82.1	32.1	12.6
0.440	69	161	34	28	78.8	71.1	82.6	33.0	14.8
0.460	69	165	30	28	80.1	71.1	84.6	30.3	14.5
0.480	69	166	29	28	80.5	71.1	85.1	29.6	14.4
0.500	69	169	26	28	81.5	71.1	86.7	27.4	14.2
0.520	65	174	21	32	81.8	67.0	89.2	24.4	15.5
0.540	61	174	21	36	80.5	62.9	89.2	25.6	17.1
0.560	61	175	20	36	80.8	62.9	89.7	24.7	17.1
0.580	56	177	18	41	79.8	57.7	90.8	24.3	18.8
0.600	56	179	16	41	80.5	57.7	91.8	22.2	18.6
0.620	55	179	16	42	80.1	56.7	91.8	22.5	19.0
0.640	54	183	12	43	81.2	55.7	93.8	18.2	19.0
0.660	54	183	12	43	81.2	55.7	93.8	18.2	19.0
0.680	53	183	12	44	80.8	54.6	93.8	18.5	19.4
0.700	48	184	11	49	79.5	49.5	94.4	18.6	21.0
0.720	43	187	8	54	78.8	44.3	95.9	15.7	22.4
0.740	43	188	7	54	79.1	44.3	96.4	14.0	22.3
0.760	39	188	7	58	77.7	40.2	96.4	15.2	23.6
0.780	36	188	7	61	76.7	37.1	96.4	16.3	24.5
0.800	35	190	5	62	77.1	36.1	97.4	12.5	24.6
0.820	30	191	4	67	75.7	30.9	97.9	11.8	26.0

0.840	26	191	4	71	74.3	26.8	97.9	13.3	27.1
0.860	23	192	3	74	73.6	23.7	98.5	11.5	27.8
0.880	21	192	3	76	72.9	21.6	98.5	12.5	28.4
0.900	16	192	3	81	71.2	16.5	98.5	15.8	29.7
0.920	14	192	3	83	70.5	14.4	98.5	17.6	30.2
0.940	8	192	3	89	68.5	8.2	98.5	27.3	31.7
0.960	4	194	1	93	67.8	4.1	99.5	20.0	32.4
0.980	2	195	0	95	67.5	2.1	100.0	0.0	32.8
1.000	0	195	0	97	66.8	0.0	100.0	.	33.2





Fit Statistics for SCORE Data

Data Set	Total Frequency	Log Likelihood	Error Rate	AIC	AICC	BIC	SC	R-Square	Max-Rescaled R-Square	AUC	Brier Score
WORK.VST	7040	-1855.7	0.1163	3723.414	3723.426	3764.57	3764.57	-0.46485	-3.43427	0.905631	0.080897

I.3.2. TSV-S

Input

```
title 'logistic analysis for small TSVZ2 -- backward';
proc logistic data=work.sm_TSVZ2 outest=betas covout plots=(roc);
class lgveh frtpax wt_class;
```

```
model der(event="1")=  trnspd2|wt_class|totleng|lgveh / selection=backward
waldcl clparm=wald ctable lackfit expb outroc=roc1;
output out=pred predprobs=(individual crossvalidate);
score data=work.TSVZ fitstat;
run ;;
```

Output

logistic analysis for small TSVZ2 -- backward

The LOGISTIC Procedure

Model Information

Data Set	WORK.SM_TSVZ2	
Response Variable	DER	DER
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	179
Number of Observations Used	179

Response Profile

Ordered Value	DER	Total Frequency
1	0	120
2	1	59

Probability modeled is DER=1.

Backward Elimination Procedure

Class Level Information

Class	Value	Design Variables
LGVEH	**N**	1
	Y	-1
wt_class	**FRT_LOCO**	1
	PAX_LOCO	-1

Step 0. The following effects were entered:

Intercept TRNSPD2 wt_class TRNSPD2*wt_class TOTLENG TRNSPD2*TOTLENG TOTLENG*wt_class TRNSPD*TOTLEN*wt_cla LGVEH TRNSPD2*LGVEH LGVEH*wt_class TRNSPD*LGVEH*wt_clas TOTLENG*LGVEH TRNSPD*TOTLENG*LGVEH TOTLEN*LGVEH*wt_clas TRNS*TOTL*LGVE*wt_cl

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	166.901
SC	232.124	217.899
-2 Log L	226.937	134.901

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	92.0354	15	<.0001
Score	76.3803	15	<.0001
Wald	38.1914	15	0.0008

Step 1. Effect TRNS*TOTL*LGVE*wt_cl is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	165.025
SC	232.124	212.836
-2 Log L	226.937	135.025

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	91.9113	14	<.0001
Score	75.9207	14	<.0001
Wald	38.1054	14	0.0005

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.1345	1	0.7138

Step 2. Effect TRNSPD*LGVEH*wt_clas is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	163.043
SC	232.124	207.667
-2 Log L	226.937	135.043

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	91.8932	13	<.0001
Score	75.8217	13	<.0001
Wald	38.1099	13	0.0003

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.1226	2	0.9405

Step 3. Effect TOTLEN*LGVEH*wt_clas is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	161.524
SC	232.124	202.960
-2 Log L	226.937	135.524

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	91.4128	12	<.0001
Score	75.7914	12	<.0001
Wald	38.5853	12	0.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.4789	3	0.9235

Step 4. Effect TRNSPD*TOTLEN*wt_cla is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	159.923
SC	232.124	198.172
-2 Log L	226.937	135.923

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	91.0135	11	<.0001
Score	75.7754	11	<.0001
Wald	38.5190	11	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.8954	4	0.9252

Step 5. Effect TRNSPD2*wt_class is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	158.009
SC	232.124	193.071
-2 Log L	226.937	136.009

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	90.9274	10	<.0001
Score	75.4053	10	<.0001
Wald	37.9709	10	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
1.0706	5	0.9567

Step 6. Effect TOTLENG*wt_class is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	157.471
SC	232.124	189.345
-2 Log L	226.937	137.471

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	89.4660	9	<.0001
Score	74.3612	9	<.0001
Wald	38.9802	9	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
2.3807	6	0.8816

Step 7. Effect TRNSPD*TOTLENG*LGVEH is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	158.094
SC	232.124	186.780
-2 Log L	226.937	140.094

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.8432	8	<.0001
Score	74.3058	8	<.0001
Wald	41.4633	8	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
4.8330	7	0.6803

Step 8. Effect TRNSPD2*LGVEH is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	156.097
SC	232.124	181.596
-2 Log L	226.937	140.097

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.8393	7	<.0001
Score	71.0678	7	<.0001
Wald	41.6754	7	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
4.8394	8	0.7746

Step 9. Effect TOTLENG*LGVEH is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	154.585
SC	232.124	176.896
-2 Log L	226.937	140.585

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.3519	6	<.0001
Score	70.9831	6	<.0001
Wald	42.4885	6	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
5.3678	9	0.8011

Step 10. Effect LGVEH*wt_class is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	154.911
SC	232.124	174.035
-2 Log L	226.937	142.911

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.0259	5	<.0001
Score	69.1040	5	<.0001
Wald	43.2677	5	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
7.1064	10	0.7154

Step 11. Effect wt_class is removed:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.937	153.244
SC	232.124	169.181
-2 Log L	226.937	143.244

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	83.6923	4	<.0001
Score	69.0459	4	<.0001
Wald	43.2527	4	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
7.6050	11	0.7482

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Variable Label
1	TRNS*TOTL*LGVE*wt_cl	1	14	0.1330	0.7153	
2	TRNSPD*LGVEH*wt_clas	1	13	0.0177	0.8940	

3	TOTLEN*LGVEH*wt_clas	1	12	0.3080	0.5789	
4	TRNSPD*TOTLEN*wt_cla	1	11	0.4142	0.5198	
5	TRNSPD2*wt_class	1	10	0.0875	0.7674	
6	TOTLENG*wt_class	1	9	1.2888	0.2563	
7	TRNSPD*TOTLENG*LGVEH	1	8	2.4533	0.1173	
8	TRNSPD2*LGVEH	1	7	0.0038	0.9506	
9	TOTLENG*LGVEH	1	6	0.4834	0.4869	
10	LGVEH*wt_class	1	5	1.9523	0.1623	
11	wt_class	1	4	0.3293	0.5661	wt_class

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
TRNSPD2	1	14.9569	0.0001
TOTLENG	1	6.8276	0.0090
TRNSPD2*TOTLENG	1	5.6572	0.0174
LGVEH	1	38.1410	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-5.2729	1.2119	18.9295	<.0001	0.005
TRNSPD2		1	0.0893	0.0231	14.9569	0.0001	1.093
TOTLENG		1	0.0362	0.0139	6.8276	0.0090	1.037
TRNSPD2*TOTLENG		1	-0.00075	0.000315	5.6572	0.0174	0.999
LGVEH	N	1	-1.5733	0.2548	38.1410	<.0001	0.207

Odds Ratio Estimates

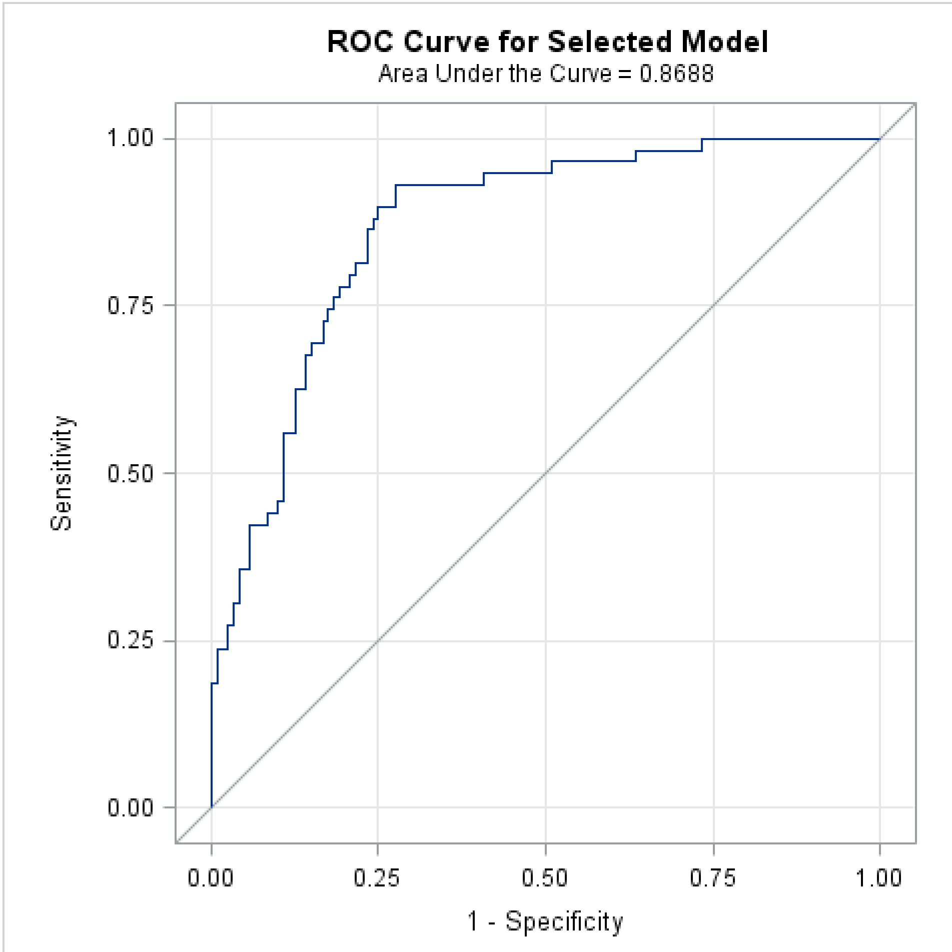
Effect	Point Estimate	95% Wald Confidence Limits	
LGVEH N vs Y	0.043	0.016	0.117

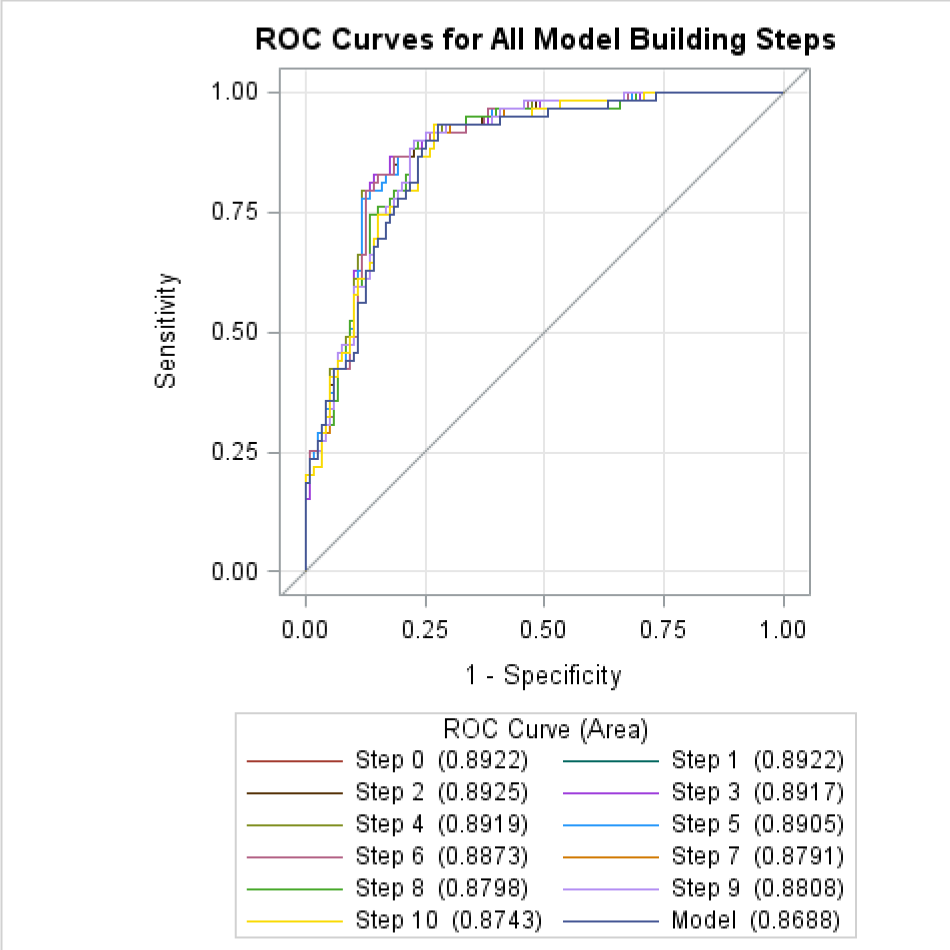
Association of Predicted Probabilities and Observed Responses

Percent Concordant	86.9	Somers' D	0.738
Percent Discordant	13.1	Gamma	0.738
Percent Tied	0.0	Tau-a	0.328
Pairs	7080	c	0.869

Parameter Estimates and Wald Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Intercept		-5.2729	-7.6483	-2.8975
TRNSPD2		0.0893	0.0440	0.1345
TOTLENG		0.0362	0.00904	0.0633
TRNSPD2*TOTLENG		-0.00075	-0.00137	-0.00013
LGVEH	N	-1.5733	-2.0726	-1.0740


ROC Curve for Selected Model
Area Under the Curve = 0.8688
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity



Partition for the Hosmer and Lemeshow Test

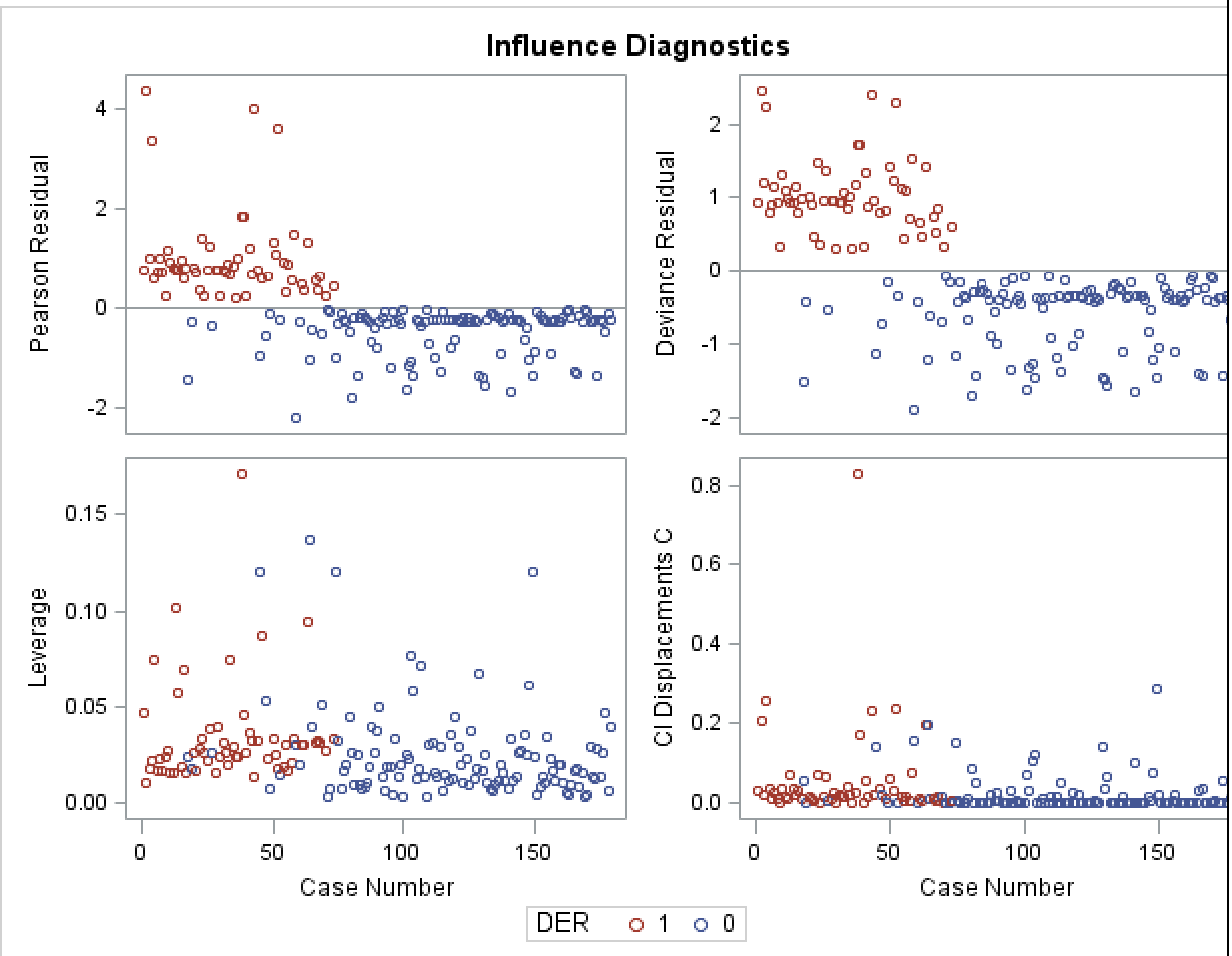
Group	Total	DER = 1		DER = 0	
		Observed	Expected	Observed	Expected
1	18	0	0.15	18	17.85
2	18	1	0.71	17	17.29
3	18	1	1.06	17	16.94
4	18	1	1.29	17	16.71
5	18	1	2.05	17	15.95
6	18	9	6.05	9	11.95
7	18	9	9.26	9	8.74
8	18	12	11.31	6	6.69
9	18	11	12.35	7	5.65
10	17	14	14.78	3	2.22

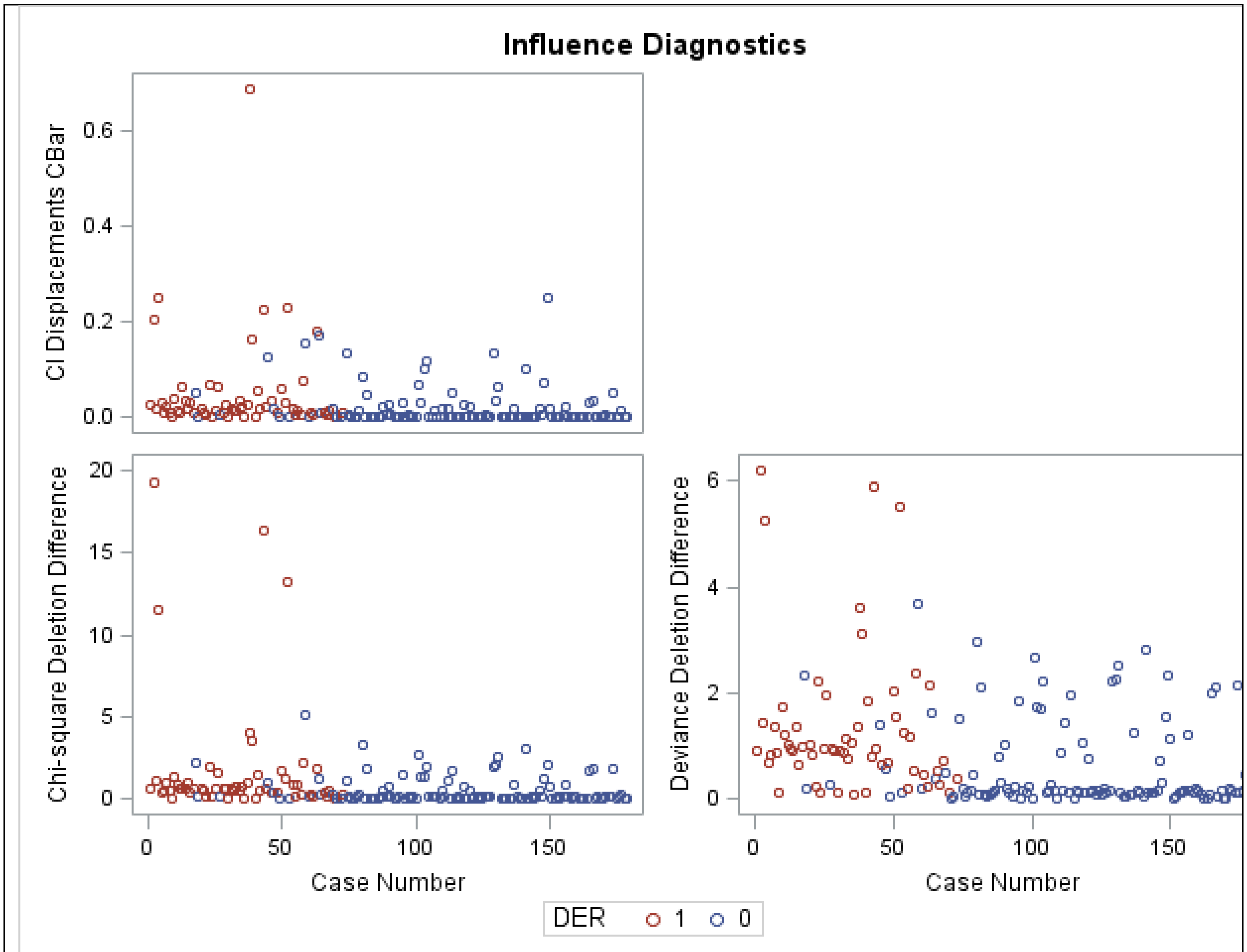
Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
4.0404	8	0.8535

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	59	0	120	0	33.0	100.0	0.0	67.0	.
0.020	59	19	101	0	43.6	100.0	15.8	63.1	0.0
0.040	59	27	93	0	48.0	100.00	22.5	61.2	0.0
0.060	56	44	76	3	55.9	94.9	36.7	57.6	6.4
0.080	55	69	51	4	69.3	93.2	57.5	48.1	5.5
0.100	55	77	43	4	73.7	93.2	64.2	43.9	4.9
0.120	54	79	41	5	74.3	91.5	65.8	43.2	6.0
0.140	54	81	39	5	75.4	91.5	67.5	41.9	5.8
0.160	54	83	37	5	76.5	91.5	69.2	40.7	5.7
0.180	54	83	37	5	76.5	91.5	69.2	40.7	5.7
0.200	53	84	36	6	76.5	89.8	70.0	40.4	6.7
0.220	53	86	34	6	77.7	89.8	71.7	39.1	6.5
0.240	53	87	33	6	78.2	89.8	72.5	38.4	6.5
0.260	53	88	32	6	78.8	89.8	73.3	37.6	6.4
0.280	53	88	32	6	78.8	89.8	73.3	37.6	6.4
0.300	52	88	32	7	78.2	88.1	73.3	38.1	7.4
0.320	51	89	31	8	78.2	86.4	74.2	37.8	8.2
0.340	50	91	29	9	78.8	84.7	75.8	36.7	9.0
0.360	49	92	28	10	78.8	83.1	76.7	36.4	9.8
0.380	48	92	28	11	78.2	81.4	76.7	36.8	10.7
0.400	47	92	28	12	77.7	79.7	76.7	37.3	11.5
0.420	46	94	26	13	78.2	78.0	78.3	36.1	12.1
0.440	46	94	26	13	78.2	78.0	78.3	36.1	12.1
0.460	45	95	25	14	78.2	76.3	79.2	35.7	12.8
0.480	45	97	23	14	79.3	76.3	80.8	33.8	12.6
0.500	42	97	23	17	77.7	71.2	80.8	35.4	14.9
0.520	41	98	22	18	77.7	69.5	81.7	34.9	15.5
0.540	40	99	21	19	77.7	67.8	82.5	34.4	16.1
0.560	38	100	20	21	77.1	64.4	83.3	34.5	17.4
0.580	37	101	19	22	77.1	62.7	84.2	33.9	17.9
0.600	35	104	16	24	77.7	59.3	86.7	31.4	18.8
0.620	31	105	15	28	76.0	52.5	87.5	32.6	21.1
0.640	25	107	13	34	73.7	42.4	89.2	34.2	24.1
0.660	23	108	12	36	73.2	39.0	90.0	34.3	25.0
0.680	19	111	9	40	72.6	32.2	92.5	32.1	26.5
0.700	18	114	6	41	73.7	30.5	95.0	25.0	26.5
0.720	14	114	6	45	71.5	23.7	95.0	30.0	28.3
0.740	14	116	4	45	72.6	23.7	96.7	22.2	28.0
0.760	13	117	3	46	72.6	22.0	97.5	18.8	28.2
0.780	12	118	2	47	72.6	20.3	98.3	14.3	28.5
0.800	12	119	1	47	73.2	20.3	99.2	7.7	28.3
0.820	11	119	1	48	72.6	18.6	99.2	8.3	28.7

0.840	10	119	1	49	72.1	16.9	99.2	9.1	29.2
0.860	10	120	0	49	72.6	16.9	100.0	0.0	29.0
0.880	9	120	0	50	72.1	15.3	100.0	0.0	29.4
0.900	7	120	0	52	70.9	11.9	100.0	0.0	30.2
0.920	6	120	0	53	70.4	10.2	100.0	0.0	30.6
0.940	5	120	0	54	69.8	8.5	100.0	0.0	31.0
0.960	0	120	0	59	67.0	0.0	100.0	.	33.0





Fit Statistics for SCORE Data

Data Set	Total Frequency	Log Likelihood	Error Rate	AIC	AICC	BIC	SC	R-Square	Max-Rescaled R-Square	AUC	Brier Score
WORK.TSVZ	11308	-3068.5	0.1666	6146.904	6146.909	6183.57	6183.57	-0.61047	-9.53312	0.879029	0.087788

I.3.3. TSV-M

Input

```
title 'logistic analysis for small TSVNZ incl totleng stepwise';
proc logistic data=work.sm_TSVNZ2 outest=betas covout plots=(roc);
class lgveh frtpax wt_class;
```

```
model der(event="1")=  vehspd2|trnspd2|wt_class|totleng|lgveh/
selection=stepwise waldcl clparm=wald ctable lackfit expb outroc=roc1;
output out=pred predprobs=(individual crossvalidate);
score data=work.TSVNZ fitstat;
run ;;
```

Output

logistic analysis for small TSVNZ incl totleng stepwise

The LOGISTIC Procedure

Model Information

Data Set	WORK.SM_TSVNZ2	
Response Variable	DER	DER
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	332
Number of Observations Used	332

Response Profile

Ordered Value	DER	Total Frequency
1	0	218
2	1	114

Probability modeled is DER=1.

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables	
LGVEH	**N**	1	
	Y	-1	
wt_class	**FRT_CAR**	1	0
	FRT_LOCO	0	1
	PAX_LOCO	-1	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L = 427.117

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
131.6820	32	<.0001

Step 1. Effect LGVEH entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	429.117	292.548
SC	432.922	300.159
-2 Log L	427.117	288.548

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	138.5681	1	<.0001
Score	110.3696	1	<.0001
Wald	38.2011	1	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
44.4819	31	0.0554

Note: No effects for the model in Step 1 are removed.

Step 2. Effect TRNSPD2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	429.117	288.327
SC	432.922	299.743
-2 Log L	427.117	282.327

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	144.7892	2	<.0001
Score	114.3376	2	<.0001

Wald	42.6567	2	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
32.8017	30	0.3312

Note: No effects for the model in Step 2 are removed.

Step 3. Effect vehspd2 entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	429.117	285.750
SC	432.922	300.970
-2 Log L	427.117	277.750

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	149.3666	3	<.0001
Score	116.8947	3	<.0001
Wald	44.8643	3	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
25.3573	29	0.6596

Note: No effects for the model in Step 3 are removed.

Note: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	**LGVEH**		1	1	110.3696		<.0001	LGVEH
2	**TRNSPD2**		1	2	6.1187		0.0134	TRNSPD2
3	**vehspd2**		1	3	4.4570		0.0348	vehspd2

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
vehspd2	1	4.3007	0.0381
TRNSPD2	1	7.3530	0.0067
LGVEH	1	38.0959	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-3.2144	0.5574	33.2572	<.0001	0.040
vehspd2		1	0.0243	0.0117	4.3007	0.0381	1.025
TRNSPD2		1	0.0233	0.00858	7.3530	0.0067	1.024
LGVEH	N	1	-2.2628	0.3666	38.0959	<.0001	0.104

Odds Ratio Estimates

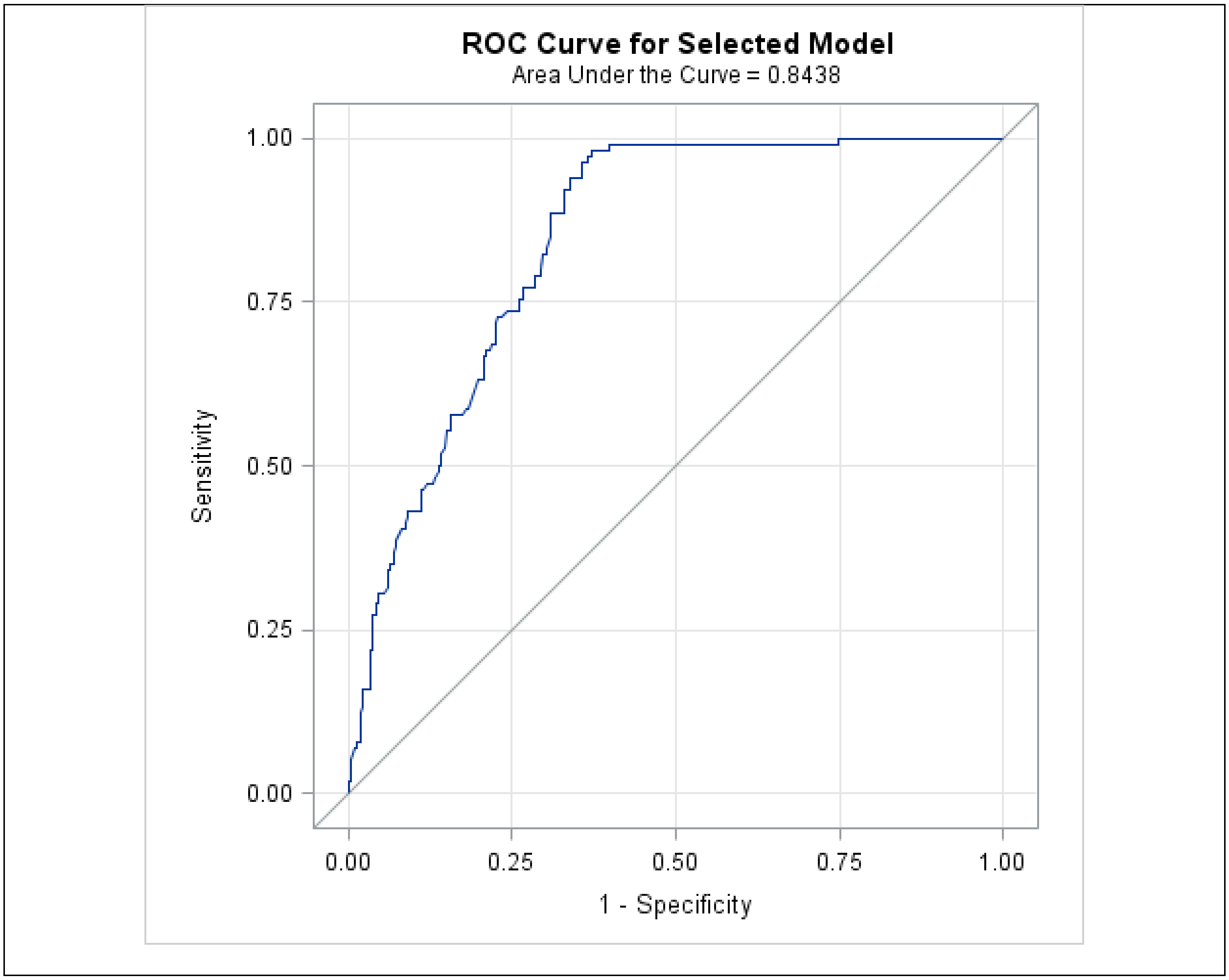
Effect	Point Estimate	95% Wald Confidence Limits	
vehspd2	1.025	1.001	1.048
TRNSPD2	1.024	1.006	1.041
LGVEH N vs Y	0.011	0.003	0.046

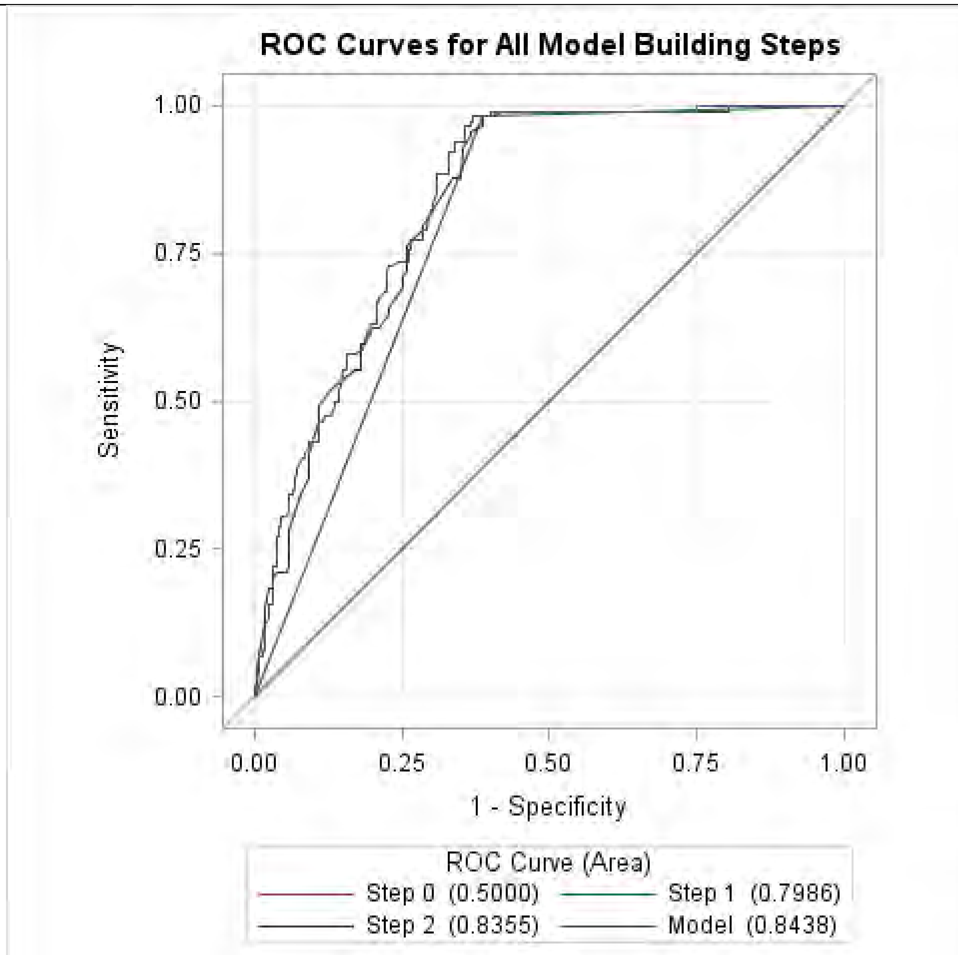
Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.3	Somers' D	0.688
Percent Discordant	15.6	Gamma	0.689
Percent Tied	0.1	Tau-a	0.311
Pairs	24852	c	0.844

Parameter Estimates and Wald Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Intercept		-3.2144	-4.3069	-2.1220
vehspd2		0.0243	0.00134	0.0473
TRNSPD2		0.0233	0.00645	0.0401
LGVEH	N	-2.2628	-2.9814	-1.5443


ROC Curve for Selected Model
Area Under the Curve = 0.8438
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity



Partition for the Hosmer and Lemeshow Test

Group	Total	DER = 1		DER = 0	
		Observed	Expected	Observed	Expected
1	33	0	0.22	33	32.78
2	33	1	0.35	32	32.65
3	33	0	0.47	33	32.53
4	33	0	0.75	33	32.25
5	33	13	12.18	20	20.82
6	33	17	15.88	16	17.12
7	33	17	17.87	16	15.13
8	33	18	19.18	15	13.82
9	34	22	21.63	12	12.37
10	34	26	25.47	8	8.53

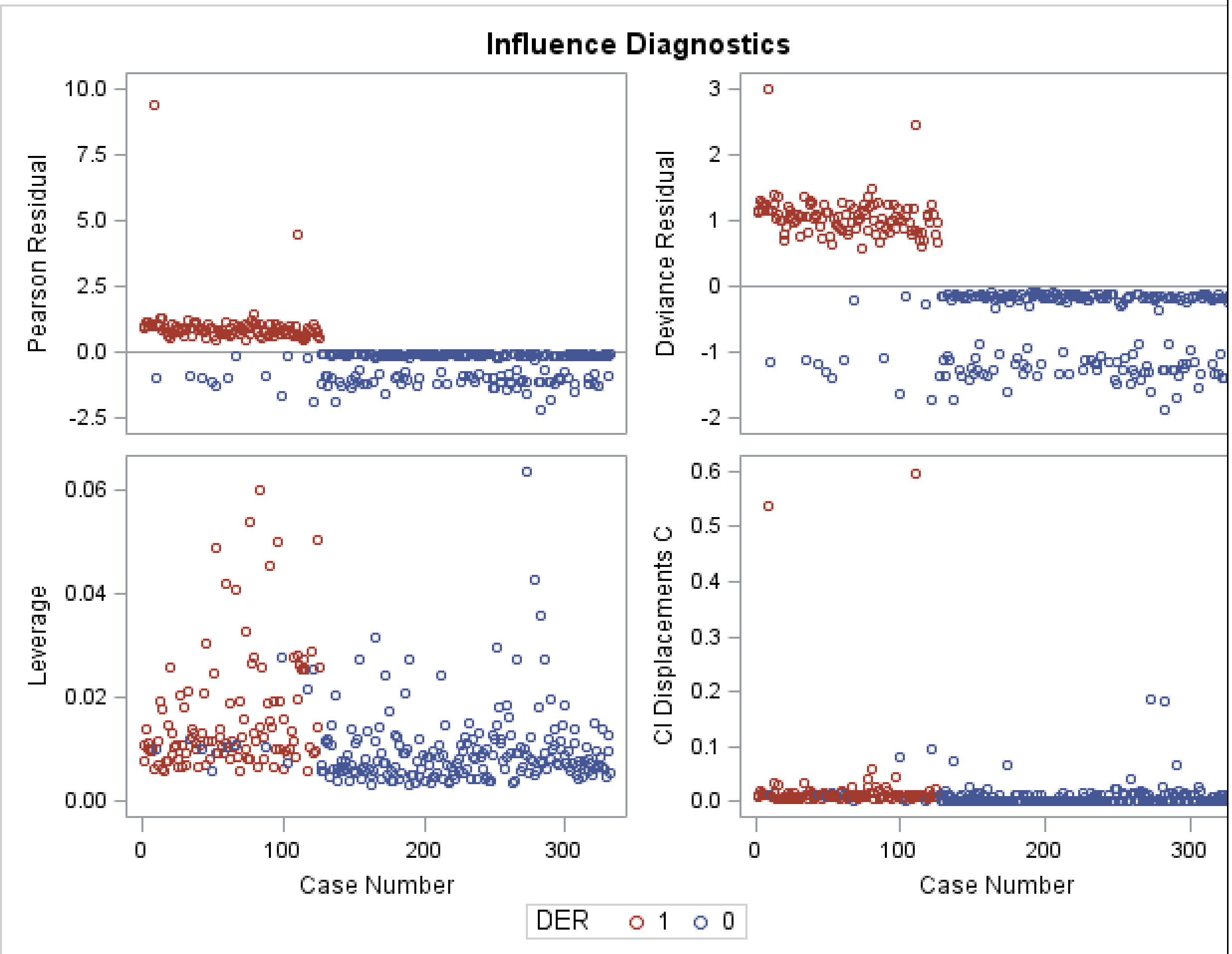
Hosmer and Lemeshow Goodness-of-Fit Test

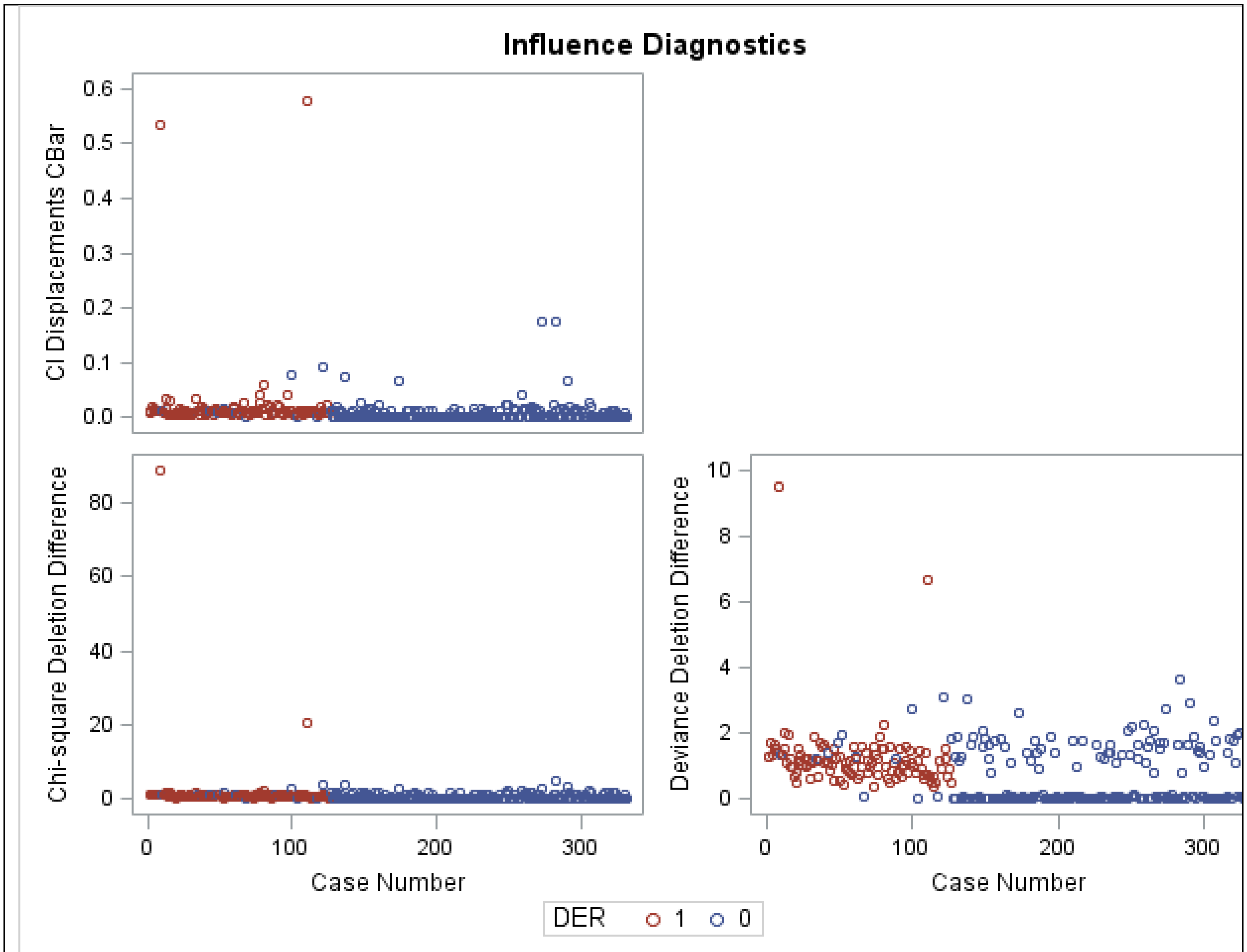
Chi-Square	DF	Pr > ChiSq
3.2856	8	0.9152

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	114	0	218	0	34.3	100.0	0.0	65.7	.
0.020	113	114	104	1	68.4	99.1	52.3	47.9	0.9
0.040	112	130	88	2	72.9	98.2	59.6	44.0	1.5
0.060	112	133	85	2	73.8	98.2	61.0	43.1	1.5
0.080	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.100	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.120	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.140	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.160	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.180	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.200	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.220	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.240	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.260	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.280	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.300	112	134	84	2	74.1	98.2	61.5	42.9	1.5
0.320	111	134	84	3	73.8	97.4	61.5	43.1	2.2
0.340	111	137	81	3	74.7	97.4	62.8	42.2	2.1
0.360	111	137	81	3	74.7	97.4	62.8	42.2	2.1
0.380	109	138	80	5	74.4	95.6	63.3	42.3	3.5
0.400	107	140	78	7	74.4	93.9	64.2	42.2	4.8
0.420	107	142	76	7	75.0	93.9	65.1	41.5	4.7
0.440	105	144	74	9	75.0	92.1	66.1	41.3	5.9
0.460	97	146	72	17	73.2	85.1	67.0	42.6	10.4
0.480	88	156	62	26	73.5	77.2	71.6	41.3	14.3
0.500	84	160	58	30	73.5	73.7	73.4	40.8	15.8
0.520	78	169	49	36	74.4	68.4	77.5	38.6	17.6
0.540	75	170	48	39	73.8	65.8	78.0	39.0	18.7
0.560	64	178	40	50	72.9	56.1	81.7	38.5	21.9
0.580	55	187	31	59	72.9	48.2	85.8	36.0	24.0
0.600	49	196	22	65	73.8	43.0	89.9	31.0	24.9
0.620	38	203	15	76	72.6	33.3	93.1	28.3	27.2
0.640	35	205	13	79	72.3	30.7	94.0	27.1	27.8
0.660	32	208	10	82	72.3	28.1	95.4	23.8	28.3
0.680	26	209	9	88	70.8	22.8	95.9	25.7	29.6
0.700	21	210	8	93	69.6	18.4	96.3	27.6	30.7
0.720	16	211	7	98	68.4	14.0	96.8	30.4	31.7
0.740	11	212	6	103	67.2	9.6	97.2	35.3	32.7
0.760	8	213	5	106	66.6	7.0	97.7	38.5	33.2
0.780	5	215	3	109	66.3	4.4	98.6	37.5	33.6
0.800	3	217	1	111	66.3	2.6	99.5	25.0	33.8
0.820	2	217	1	112	66.0	1.8	99.5	33.3	34.0

0.840	1	217	1	113	65.7	0.9	99.5	50.0	34.2
0.860	0	218	0	114	65.7	0.0	100.0	.	34.3





Fit Statistics for SCORE Data

Data Set	Total Frequency	Log Likelihood	Error Rate	AIC	AICC	BIC	SC	R-Square	Max-Rescaled R-Square	AUC	Brier Score
WORK.TSVNZ	15027	-4296.6	0.2191	8601.215	8601.218	8631.686	8631.686	-0.6204	-7.27159	0.862518	0.103843

RAIL EQUIPMENT

ACCIDENT/INCIDENT

FORM F 6180.54

ACCIDENT DOWNLOADS ON DEMAND

DATA FILE STRUCTURE
AND
FIELD INPUT SPECIFICATIONS

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
amtrak	1	A/N	1	amtrak involvement		
iyr	2 - 3	A/N	2	year of incident	5	
imo	4 - 5	A/N	2	month of incident	5	
railroad	6 - 9	A	4	railroad code (Reporting RR)	1a	
incdtno	10 - 19	A/N	10	railroad assigned number	1b	
iyr2	20 - 21	A/N	2	year of incident	5	
imo2	22 - 23	A/N	2	month of incident	5	
rr2	24 - 27	A	4	railroad code (Other RR involved)	2a	
incdtno2	28 - 37	A/N	10	other railroad assigned number	2b	
iyr3	38 - 39	A/N	2	year of incident	5	
imo3	40 - 41	A/N	2	month of incident	5	
rr3	42 - 45	A	4	railroad code (RR responsible for track maintenance)	3a	
incdtno3	46 - 55	A/N	10	RR assigned number	3b	
dummy1	56 - 59	A/N	4	blank data expansion field		
gxid	60 - 66	A/N	7	grade crossing id number	4	
year	67 - 68	A/N	2	year of accident / incident	5	
month	69 - 70	A/N	2	month of incident	5	
day	71 - 72	A/N	2	day of incident	5	
timehr	73 - 74	N	2	hour of incident	6	
timemin	75 - 76	N	2	minute of incident	6	
ampm	77 - 78	A/N	2	am or pm	6	

Indicates links, changes, or new data

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
type	79 - 80	A/N	2	type of accident: 01= derailment 02= head on collision 03= rearend collision 04= side collision 05= raking collision 06= broken train collision 07= hwy-rail crossing 08= RR Grade Crossing 09= obstruction 10= explosive – detonation 11= fire / violent rupture 12= other impacts 13= other (described in narrative)	7	
cars	81 - 83	N	3	# of cars carrying hazmat	8	
carsdmg	84 - 86	N	3	# of hazmat cars damaged or derailed	9	
carshzd	87 - 89	N	3	# of cars that released hazmat	10	
evacuate	90 - 95	N	6	# of persons evacuated	11	
division	96 - 115	A/N	20	railroad division	12	
station	116 - 135	A/N	20	nearest city and town	13	
milepost	136 - 141	A/N	6	milepost #	14	
state	142 - 143	A/N	2	FIPS State code	15	
temp	144 - 146	N	3	temperature in degrees fahrenheit	17	
visiblty	147	A/N	1	daylight period: 1=dawn 2=day 3=dusk 4=dark	18	
weather	148	A/N	1	weather conditions: 1=clear 2=cloudy 3=rain 4=fog 5=sleet 6=snow	19	
trnspd	149 - 151	A/N	3	speed of train in miles per hour: blank=unknown	28	
typspd	152	A/N	1	train speed type: E=estimated R=recorded blank=unknown	28	
trnnbr	153 - 156	A/N	4	train id number	27	
trndir	157	A/N	1	train direction: 1=north 2=south 3=east 4=west	24	
tons	158 - 162	N	5	gross tonnage, excluding power units	29	

Indicates links, changes, or new data

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
typeq	163	A/N	1	type of consist: 1=freight train 2=passenger train 3=commuter train 4=work train 5=single car 6= cut of cars 7= yard / switching 8= light loco(s) 9= maint / inspect car A= spec. MoW eq.	25	Added 'A' to selection options
eqatt	164	A/N	1	equipment attended: 1=yes 2=no	26	
trkname	165 - 184	A/N	20	track identification	21	
trkclas	185	A/N	1	FRA track class: 1-9,X	22	Extended Track class to 9
trkdnsty	186 - 191	A/N	6	annual track density - gross tonnage in millions	23	
typtrk	192	A/N	1	type of track: 1=main 2=yard 3=siding 4=industry	20	
rrcar1	193 - 196	A/N	4	car initials (first involved)	31a(1)	
carnbr1	197 - 202	A/N	6	car number (first involved)	31a(1)	
positon1	203 - 205	A/N	3	car position in train (first involved)	31b(1)	
loaded1	206	A/N	1	car loaded or not (first involved): Y=yes N=no blank=not applicable	31c(1)	
rrcar2	207 - 210	A/N	4	car initials (causing)	31a(2)	
carnbr2	211 - 216	A/N	6	car number (causing)	31a(2)	
positon2	217 - 219	A/N	3	car position in train (causing)	31b(2)	
loaded2	220	A/N	1	car loaded or not (causing): Y=yes N=no blank=not applicable	31c(2)	
headend1	221	N	1	# of head end locomotives	34a(1)	
midman1	222	N	1	# of mid train locomotives, manual	34b(1)	
midrem1	223	N	1	# of mid train locomotives, remote	34c(1)	
rman1	224	N	1	# of rear end locomotives, manual	34d(1)	
rrem1	225	N	1	# of rear end locomotives, remote	34e(1)	
headend2	226	N	1	# of head end locomotives, derailed	34a(2)	
midman2	227	N	1	# of mid train locomotives, manual, derailed	34b(2)	

Indicates links, changes, or new data

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
midrem2	228	N	1	# of mid train locomotives, remote, derailed	34c(2)	
rman2	229	N	1	# of rear end locomotives, manual, derailed	34d(2)	
rrem2	230	N	1	# of rear end locomotives, remote, derailed	34e(2)	
loadf1	231 - 233	N	3	# of loaded freight cars	35a(1)	
loadp1	234 - 236	N	3	# of loaded passenger cars	35b(1)	
emptyf1	237 - 239	N	3	# of empty freight cars	35c(1)	
emptyp1	240 - 242	N	3	# of empty passenger cars	35d(1)	
caboose1	243 - 245	N	3	# of cabooses	35e(1)	
loadf2	246 - 248	N	3	# of derailed loaded freight cars	35a(2)	
loadp2	249 - 251	N	3	# of derailed loaded passenger cars	35b(2)	
emptyf2	252 - 254	N	3	# of derailed empty freight cars	35c(2)	
emptyp2	255 - 257	N	3	# of derailed empty passenger cars	35d(2)	
caboose2	258 - 260	N	3	# of derailed cabooses	35e(2)	
eqpdmg	261 - 267	N	7	reportable equipment damage in $	36	
trkdmg	268 - 274	N	7	track, signal, way & structure damage in $	37	
cause	275 - 278	A/N	4	primary cause of incident (refer to Appendix C)	38	
cause2	279 - 282	A/N	4	contributing cause of incident (refer to Appendix C)	39	
caskldrr	283 - 285	N	3	# killed for reporting RR - calculated from Form F6180.55a's submitted		
casinjrr	286 - 289	N	4	# injured for reporting RR - calculated from Form F6180.55a's submitted		
caskld	290 - 292	N	3	total killed for all RRs involved - calculated from Form F6180.55a's submitted		
casinj	293 - 296	N	4	total injured for all RRs involved - calculated from Form F6180.55a's submitted		
accause	297 - 300	A/N	4	accident cause code from jointcd 1 record for this incident (refer to Appendix C)		
acctrk	301	A/N	1	type track code from jointcd 1 record for this incident		

Indicates links, changes, or new data

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
acctrkcl	302	A/N	1	FRA track class from jointcd 1 record for this incident (FRA track class: 1-9,X)		
highspd	303 - 305	A/N	3	maximum speed reported for equipment involved: blank=unknown		
accdmg	306 - 313	N	8	total reportable damage on all reports in $		
dummy2	314 - 316	A/N	3	blank data expansion field		
stcnty	317 - 322	A/N	6	FIPS State & County code		
totinj	323 - 326	N	4	total injured for railroad as reported on Form F6180.54		
dummy3	327 - 332	A/N	6	blank data expansion field		
totkld	333 - 336	N	4	total killed for railroad as reported on Form F6180.54		
engrs	337	A/N	1	# of engineers on duty: blank=not applicable	40	
firemen	338	A/N	1	# of firemen on duty: blank=not applicable	41	
conductr	339	A/N	1	# of conductors on duty: blank=not applicable	42	
brakemen	340	A/N	1	# of brakemen on duty: blank=not applicable	43	
enghr	341 - 342	A/N	2	# of hours engineers on duty: blank=not applicable	44	
engmin	343 - 344	A/N	2	# of minutes engineers on duty (+enghr): blank=not applicable	44	
cdtrhr	345 - 346	A/N	2	# of hours conductors on duty: blank=not applicable	45	
cdtrmin	347 - 348	A/N	2	# of minutes conductors on duty (+cdtrhr): blank=not applicable	45	
jointcd	349	A/N	1	indicates railroad reporting		
region	350	A/N	1	FRA designated region		
dummy4	351	A/N	1	blank data expansion field		
typrr	352 - 353	A/N	2	type railroad - ICC categories; 1st position indicates class 1,2,or 3 RR		

Indicates links, changes, or new data

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
dummy5	354 - 356	A/N	3	blank data expansion field		
rrdiv	357 - 362	A/N	6	RR division code		
method	363 - 382	A/N	20	Method of operation (Series of 1 position codes): A= ATCS B=auto train control C= auto train stop D= cab signals E= traffic control F= interlocking G= automatic block rules H= current of traffic I= time table / train orders J= track warrant control K= direct traffic Control L= yard limits M= special instructions N= other than main track O= other (specify in a narrative) P= positive train control	30	Added 'P' to selection options
narrlen	383 - 386	N	4	length of narrative		
dummy6	387 - 390	A/N	4	blank data expansion field		
year4	391 - 394	A/N	4	Four character year identification		
rrempkld	395 - 397	N	3	# of RR employees killed as reported on Form F6180.54	46	
rrempinj	398 - 400	N	3	# of RR employees injured as reported on Form F6180.54	46	
passkld	401 - 403	N	3	# of passengers killed as reported on Form F6180.54	47	
passinj	404 - 406	N	3	# of passengers injured as reported on Form F6180.54	47	
otherkld	407 - 409	N	3	# of others killed as reported on Form F6180.54	48	
otherinj	410 - 412	N	3	# of others injured as reported on Form F6180.54	48	
county	413 - 432	A/N	20	County Name (See FIPS Codes for associated codes)	16	
cntycd	433 - 435	A/N	3	FIPS County Code		
alcohol	436 - 437	A/N	2	# of positive alcohol tests	32	
drug	438 - 439	A/N	2	# of positive drug tests	32	
dummy7	440 - 451	A/N	12	blank data expansion field		
passtrn	452	A/N	1	were there passengers being transported: Y=yes N=no blank=not applicable	33	
ssb1	453 - 472	A/N	20	special study block 1	49	

Indicates links, changes, or new data

Accident Data On Demand

Rail Equipment Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK # ON FORM 6180.54	CONVERSION
ssb2	473 - 492	A/N	20	special study block 2	49	
narr1	493 - 592	A/N	100	narrative	52	
narr2	593 - 692	A/N	100	narrative	52	
narr3	693 - 792	A/N	100	narrative	52	
narr4	793 - 892	A/N	100	narrative	52	
narr5	893 - 992	A/N	100	narrative	52	
narr6	993 - 1092	A/N	100	narrative	52	
narr7	1093 - 1192	A/N	100	narrative	52	
narr8	1193 - 1292	A/N	100	narrative	52	
narr9	1293 - 1392	A/N	100	narrative	52	
narr10	1393 - 1492	A/N	100	narrative	52	
narr11	1493 - 1592	A/N	100	narrative	52	
narr12	1593 - 1692	A/N	100	narrative	52	
narr13	1693 - 1792	A/N	100	narrative	52	
narr14	1793 - 1892	A/N	100	narrative	52	
narr15	1893 - 1992	A/N	100	narrative	52	
rcl	1993 - 1993	A/N	1	Remote control locomotive = 0,1,2, or 3 0= not a remotely controlled operation 1= remote control portable transmitter 2= remote control tower operation 3= remote control portable transmitter (more than one remote control)	30a	new data
latitude	1994 - 2003	N	10	Latitude in decimal degrees, explicit decimal, explicit +/- (WGS84)	50	new data
longitud	2004 - 2014	N	11	Longitude in decimal degrees, explicit decimal, explicit +/- (WGS84)	51	new data

Indicates links, changes, or new data

HIGHWAY-RAIL GRADE CROSSING

ACCIDENT/INCIDENT

FORM F 6180.57

ACCIDENT DOWNLOADS ON DEMAND

DATA FILE STRUCTURE

AND

FIELD INPUT SPECIFICATIONS

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
amtrak	1	A/N	1	amtrak involvement		
iyr	2 - 3	A/N	2	year of incident	5	
imo	4 - 5	A/N	2	month of incident	5	
railroad	6 - 9	A	4	railroad code (Reporting RR)	1a	
incdtno	10 - 19	A/N	10	railroad assigned number	1b	
iyr2	20 - 21	A/N	2	year of incident	5	
imo2	22 - 23	A/N	2	month of incident	5	
rr2	24 - 27	A	4	railroad code (Other RR involved)	2a	
incdtno2	28 - 37	A/N	10	other railroad assigned number	2b	
iyr3	38 - 39	A/N	2	year of incident	5	
imo3	40 - 41	A/N	2	month of incident	5	
rr3	42 - 45	A	4	railroad code (RR responsible for track maintenance)	3a	
incdtno3	46 - 55	A/N	10	RR assigned number	3b	
dummy1	56	A/N	1	blank data expansion field		
casinjrr	57 - 59	N	3	# of injured for reporting Railroad calculated from F6180.55a's submitted		
gxid	60 - 66	A/N	7	grade crossing id number	4	
year	67 - 68	A/N	2	year of incident	5	
month	69 - 70	A/N	2	month of incident	5	
day	71 - 72	A/N	2	day of incident	5	

Indicates links, changes, or new data

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
timehr	73 - 74	N	2	hour of incident	6	
timemin	75 - 76	N	2	minute of incident	6	
ampm	77 - 78	A/N	2	am or pm	6	
station	79 - 98	A/N	20	Nearest Timetable Station	7	
county	99 - 118	A/N	20	County Name (see FIPS Codes for associated code)	9	
state	119 - 120	A/N	2	FIPS State Code	10	
region	121	A/N	1	FRA designated region		
dummy2	122	A/N	1	blank data expansion field		
city	123 - 142	A/N	20	City name (see FIPS Codes for associated code)	11	
highway	143 - 162	A/N	20	highway name	12	
vehspd	163 - 165	A/N	3	vehicle estimated speed: blank-Unknown	14	
typveh	166	A/N	1	highway user: A= auto B= truck C= truck-trailer D= pick-up truck E= van F= bus G= school bus H= motorcycle J= other motor veh. K=pedestrian M=other	13	
vehdir	167	A/N	1	highway user direction 1=north 2=south 3=east 4=west	15	
position	168	A/N	1	position of highway user: 1=stalled on crossing 2=stopped on crossing 3=moving over crossing 4=trapped	16	

Indicates links, changes, or new data

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
rrequip	169	A/N	1	RR equipment involved: 1= train(units pulling) 2= train (units pushing) 3= train(standing) 4=cars moving 5= car/s(standing 6= light loco/s(moving) 7= light loco/s (standing) 8= other A= training pulling (RCL) B= train pushing (RCL) C= train standing (RCL)	17	Added new equipment
rrcar	170 - 172	A/N	3	position of car unit in train	18	
typacc	173	A/N	1	circumstance of accident: 1=rail equipment struck highway user 2=rail equipment struck by highway user	19	
hazard	174	A/N	1	entity transporting hazmat: 1=highway user 2=rail equipment 3=both 4=neither	20a	
temp	175 - 177	N	3	temperature in degrees Fahrenheit	21	
visiblty	178	A/N	1	Visibility: 1=dawn 2=day 3=dusk 4=dark	22	
weather	179	A/N	1	weather conditions: 1=clear 2=cloudy 3=rain 4=fog 5=sleet 6=snow	23	
typeq	180	A/N	1	type of consist: 1= freight train 2= passenger train 3= commuter train 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	24	added new selection option

Indicates links, changes, or new data

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
typtrk	181	A/N	1	type of track: 1=main 3= siding 2= yard 4= industry	25	
trkname	182 - 201	A/N	20	track identification	26	
trkclas	202	A/N	1	FRA track class: 1-9,X	27	Increased FRA track classes from 6 to 9
nbrlocos	203 - 204	N	2	number of locomotive units	28	
nbrcars	205 - 207	N	3	number of cars	29	
trnspd	208 - 210	A/N	3	speed of train in miles per hour blank=unknown	30	
typspd	211	A/N	1	train speed type: E=estimated, R=recorder, blank=unknown	30	
trndir	212	A/N	1	time table direction 1=north, 2=south, 3=east, 4=west	31	
signal	213	A/N	1	type of signaled crossing warning: if block 32 (crossing) = 01-06, then signal = 1-7 (see back of form 57 for valid entries)	33	
locwarn	214	A/N	1	location of warning: 1=both sides 2=side of vehicle approach 3=opposite side of vehicle approach	35	
warnsig	215	A/N	1	crossing warning interconnected with highway signals: 1=yes 2=no 3=unknown	36	
lights	216	A/N	1	lights at crossing: 1=yes 2=no 3=unknown	37	
standveh	217	A/N	1	motorist passed highway standing vehicle: 1=yes 2=no 3=unknown	42	
train2	218	A/N	1	motorist struck or was struck by 2nd train: 1=yes 2=no 3=unknown	40	

Indicates links, changes, or new data

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
motorist	219	A/N	1	action of motorist: 1=drove around or thru the gate 2=stopped and then proceeded 3=did not stop 4=stopped on crossing 5=other	41	
view	220	A/N	1	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	43	
vehdmg	221 - 226	N	6	highway vehicle property damage in $	47	
driver	227	A/N	1	highway vehicle driver casualty: 1=killed 2=injured 3=uninjured	44	
inveh	228	A/N	1	highway driver in vehicle: 1=yes 2=no	45	
totkld	229 - 232	N	4	total killed for railroad as reported on F6180.57		
totinj	233 - 236	N	4	total injured for railroad as reported on F6180.57		
totocc	237 - 240	N	4	total # in highway vehicle	48	
incdrpt	241	A/N	1	F6180.54 filed: 1=yes 2=no	51	
jointcd	242	A/N	1	indicates railroad reporting		
typrr	243 - 244	A/N	2	type railroad - ICC categories; 1st position indicates class 1,2,or 3 RR		
dummy3	245	A/N	1	blank data expansion field		
caskldrr	246 - 247	N	2	# killed for reporting RR - calculated F6180.55a's submitted		
dummy4	248 - 249	A/N	2	blank data expansion field		

Indicates links, changes, or new data

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
crossing	250 - 273	A/N	24	type of warning device at crossing (series of 2 digit codes) 01= gates 02= cantilever fls 03= standard fls 04= wig wags 05= highway traffic Signals 06= audible 07=cross bucks 08=stop signs 09=watchman 10=flagged by crew 11=other (specify) 12= none	32	
narrlen	274 - 277	N	4	length of narrative		
dummy5	278 - 281	A/N	4	blank data expansion field		
year4	282 - 285	A/N	4	4 digit year of incident	5	
division	286 - 305	A/N	20	railroad division	8	
public	306	A/N	1	public crossing: 1=public 2=private	12	
cntycd	307 - 309	A/N	3	FIPS county code		
stcnty	310 - 315	A/N	6	FIPS state and county code		
hzmrlsed	316	A/N	1	hazmat released by: 1=highway user 3=both 2=rail equipment 4=neither blank=unknown	20b	
hzmname	317 - 346	A/N	30	name of hazmat released	20c	
hzmqnty	347 - 349	A/N	3	quantity of hazmat released	20c	
hzmmeas	350 - 353	A/N	4	measure used in hazmat quantity field	20c	
sigwarnx	354	A/N	1	further definition of signal field: if signal=5-7, then sigwarnx=A-S (see back of form 57 for valid entries)	33	
whisban	355	A/N	1	whistle ban in effect: 1=yes 2=no 3=not provided blank=unknown	34	
drivage	356 - 357	A/N	2	vehicle driver's age: blank=unknown	38	

Indicates links, changes, or new data

Accident Data On Demand

Rail Grade Crossing Accident/Incident
Accident Downloads on Demand
Data File Structure and Field Input Specifications

FIELD NAME	FILE POSITION	FIELD TYPE	FIELD LENGTH	DEFINITION	BLOCK# ON FORM 6180.57	CONVERSION
drivgen	358	A/N	1	vehicle driver's gender: 1=male 2=female blank=unknown	39	
pleontrn	359 - 362	A/N	4	total # of people on train (includes passengers and crew): blank=unknown	50	
ssb1	363 - 382	A/N	20	special study block 1	53a	
ssb2	383 - 402	A/N	20	special study block 2	53b	
userkld	403 - 405	N	3	# of highway-rail crossing users killed as reported by railroad on F6180.57	46	
userinj	406 - 408	N	3	# of highway-rail crossing users injured as reported by railroad on F6180.57	46	
rrempkld	409 - 411	N	3	# of highway employees killed as reported by railroad on F6180.57	49	
rrempinj	412 - 414	N	3	# of railroad employees injured as reported by railroad on F6180.57	49	
passkld	415 - 417	N	3	# of train passengers killed as reported by railroad on F6180.57	52	
passinj	418 - 420	N	3	# of train passengers injured as reported by railroad on F6180.57	52	
narr1	421- 520	A/N	100	narrative	54	
narr2	521 - 620	A/N	100	narrative	54	
narr3	621 - 720	A/N	100	narrative	54	
narr4	721 - 820	A/N	100	narrative	54	
narr5	821 - 920	A/N	100	narrative	54	

Indicates links, changes, or new data

U.S. DOT CROSSING INVENTORY FORM

DATA FILE STRUCTURE

AND

FIELD INPUT SPECIFICATIONS

FORMAT FOR FRA INVENTORY FIELDS

DATA ENTRY FIELD DESCRIPTION

INVENTORY FIELD ORDER

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)		DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
1	B.	CROSSING	Crossing No.	7 C	1	(7)	Valid Crossing I.D. No. Must be 6 numeric characters followed by 1 alphabetic character.
2	D.	EFFDATE	Effective Date	6 C	8	(13)	Entered in form as MM/DD/YYYY (stored in EFFDATE field as YYMMDD)
3		EDATE	End Date	6 C	14	(19)	End date for the most current record is always '999999'. When the crossing is updated with a new record, the end date of the previous current record is set to one day before the effective date of the new current record. EDATE is stored as YYMMDD.
4	C.	REASON	Reason for Update	1 C	20		1=Changes in Existing Crossing Data 2=New Crossing 3=Closed Crossing or Abandoned
5	I.2.	STATE	State	2 C	21	(22)	Use 2-character state code. Click here to go to Valid State FIPS Code .
6	I.3.	CNTYCD	County	4 C	23	(26)	Use 4-character county code. Click here to go to Valid County FIPS Code
7		STATE2	State	2 C	27	(28)	Use 2-character state code. Click here to go to Valid State FIPS Code
8	I.12.	CITYCD	City	4 C	29	(32)	Use 4-character city code. Click here to go to Valid City FIPS Code
9	I.12.	NEAREST	In or Near City	1 C	33		0 = In City 1=Near City
10	I.1.	RAILROAD	Railroad Operating Company	4 C	34	(37)	Valid Railroad Code For valid railroad codes, refer to current list of railroad codes provided by FRA Office of Safety
11	I.4.	RRDIV	RR Division	14 C	38	(51)	Railroad Division Name or Blank
12	I.5.	RRSUBDIV	RR Subdivision	14 C	52	(65)	Railroad Subdivision or Blank
13	I.14.	HIGHWAY	Highway type and No.	7 C	66	(72)	Any Alphanumeric Data or Blank
14	I.13.	STREET	Street or Road Name	17 C	73	(89)	Any Alphanumeric Data or Blank
15	I.8.	RRID	RR I.D. No.	10 C	90	(99)	
16	I.9.	TTSTN	Nearest RR Timetable Station	6 C	100	(105)	Valid Timetable Station
17	I.6.	BRANCH	Branch or Line Name	15 C	106	(120)	Branch/Line Name or Blank

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER
(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)		DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
18	I.7.	MILEPOST	RR Milepost	6 C	121	(126)	The first two spaces can be alphanumeric, and the next four spaces numeric. There is an implied decimal point after the first 4 characters.
19	I.22.	MAPREF	County Map Ref. No.	10 C	127	(136)	Any Alphanumeric Data or Blank
20	I.17	TYPEXING	Type of Crossing	1 C	137		1=Pedestrian, 2=Private Vehicle, 3=Public Vehicle (The following is the key for the crossing type and position: 11 - Pedestrian at grade 12 - Pedestrian RR under 13 - Pedestrian RR over 21 - Private at grade 22 - Private RR under 23 - Private RR over 31 - Public at grade 32 - Public RR under 33 - Public RR over)
21	I.18.	POSXING	Position of Crossing	1 C	138		1=At grade under 2=RR Under 3=RR over
22	I.27.A	PRVCAT	Private Xing Category	1 C	139		1=Farm 2=Residential 3=Recreational 4=Industrial 5=Commercial
23	I.27.C.	PRVIND	Signs/ Signals	1 C	140		Current Values: 1=signs 2=signals 3=no signs or signals 4=both signs and signals On Previous Version of Inventory Form: 8=Signs 9=Signals 0=None
24	I.27.C.	PRVSIGN	Signs-Specify	15 C	141	(155)	Any Alphanumeric Data (Reference Field 140, PRVSIGNL)
25	A.	INIT	Initiating Agency	1 C	156		1. =Railroad 2. =State 3.=DOT 4. =Original FRA internal use. Note: 3 & 4 are for internal FRA use only.
26		BATCH	System coded Field	6 C	157	(162)	Coded field, which is used for batch identification during update: The first character is the last character of the year; The second-fourth characters are the day of the year, and the fifth-sixth characters are the sequence number.
27		USERCD		1 C	163		This field is not currently used
28		UPDATE		2 C	164	(165)	No Longer Used Previous: Coded date of update. Refer to field 105 (UPDATDAT)

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER
(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
29		LINK		5 C	166 (170)	Not in use. Previous Value: 1.Used for High Speed Corridor. 2.This was the link identification code (LIC) from the rail network model for the line on which the crossing lies. The LIC is a five-digit code incorporating the alphabetical abbreviation of the owning railroad and a sequence number. Refer to field 89 (HSCORRID)
30	II.1.C.	DAYTHRU	Day Thru Train Movements	2 N	171 (172)	0 to 99
31		DAYSWT	Switching	2 N	173 (174)	(Previous Values: 0 to 99) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field135, TOTALTRN, and Field 136 TOTALSWT)
32		NGHTTHRU	Night Thru Train Movements	2 N	175 (176)	(Previous Values: 0 to 99) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field 135, TOTALTRN)
33		NGHTSWT	Night Switching Movements	2 N	177 (178)	(Previous Values: 0 to 99) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field 135, TOTALTRN, and Field 136 TOTALSWT)
34	II.1.D.	LT1MOV	Less Than One Movement Per Day?	1 C	179	0 = At least one train per day 1= Less than one train per day Enter a check if train frequency is less than one train per day.
35	II.2.A.	MAXTTSPD	Maximum Timetable Speed	3 N	180 (182)	Values are 1 to 150
36	II.2.B	MINSPD	From Min:	3 N	183 (185)	Values are 1 to 150
37	II.2.B.	MAXSPD	To Max:	3 N	186 (188)	Values are 1 to 150
38	II.3.	MAINTRK	Main	1 N	189	Values are 0 to 9 for main track
39	II.3.	OTHRTRK	Other	2 N	190 (191)	Values are 0 to 99 for other tracks
40	II.3.	OTHRDES	Specify	10 C	192 (201)	Description, if other tracks exist

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION

INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
41	II.4.	SEPIND	Does Another RR Operate a Separate Trk. (Y/N)?	1 C	202	1=Yes 2=No
42	II.4.	SEPRR	Specify	16 C	203 (218)	Up to 4 valid RR codes Code should not be repeated
43	II.5.	SAMEIND	Does Another RR Operate Over Your Trk. (Y/N)?	1 C	219	1=Yes 2=No
44	II.5	SAMERR	Specify	16 C	220 (235)	Up to 4 valid RR codes Code should not be repeated
45		WDCODE	Warning Device Code	1 C	236	Highway warning device class at crossing. New Values: 1 - No signs or signals 2 - Other signs or signals 3 - Crossbucks 4 - Stop signs 5 - Special Active Warning Devices 6 - Highway traffic signals, wigwags, bells, or other activated 7. Flashing lights 8 - All other Gates 9 - Four Quad (full barrier) Gates (Note: SPECPRO (Field 64) has WDCODE=6; and WARNACTO (Field 142) has WDCODE=6).: Previous Values 1 - No sign or signal 2 - Other signs or signals 3 - Stop signs 4 - Crossbucks 5 - Non-train activated special protection 6 - Highway traffic signals, wigwags, or bells 7 - Flashing lights 8 - Gates
46		XBUCKRF	Crossbucks-Reflectorized	1 N	237	(Previous Values: 0 to 9) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field 138, XBUCK)
47		XBUCKNRF	Crossbucks-Non-reflectorized	1 N	238	(Previous Values: 0 to 9) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field 138, XBUCK)

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
48	III.2.B.	STOPSTD	Highway Stop Signs	1 N	239	0 to 9 9 represents 9 or more
49		STOPOTH	Other Stop Sign	1 N	240	Previous Values: (0 to 9, 9 represents 9 or more) Not in New Form-field No Longer Maintained in Inventory-obsolete Conversion: If at least one of the two "Other Signs: Specify" field sets (OTHSGN1 and OTHDES1, or OTHSGN2 and OTHDES2) are blank, the value for STOPOTH (Other Stop Sign) was placed in the blank OTHSGN1 (or OTHSGN2) field, and "OTHRSTPSGN" was entered in the corresponding OTHDES1 (or OTHDES2) field.
50	III.2.F.	OTHSGN1	Other Signs:	1 N	241	0 to 9 9 represents 9 or more
51	III.2.F.	OTHDES1	Specify:	10 C	242 (251)	Any Alphanumeric Description
52	III.2.F.	OTHSGN2	Other Signs:	1 N	252	0 to 9 9 represents 9 or more
53	III.2.F.	OTHDES2	Specify:	10 C	253 (262)	Any Alphanumeric Description
54		GATERW	Gates-Red & White	1 N	263	Previous Values: 0 to 9 (9 represents 9 or more) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field 139, GATES)
55		GATEOTH	Gates-Other	1 N	264	(Previous Values: 0 to 9, (9 represents 9 or more) Not in New Form-field No Longer Maintained in Inventory-obsolete (Reference Field 139, GATES)
56	III.3.C.	FLASHOV	Canti-levered (or bridged) Flashing Lights-Over Traffic Lane	1 N	265	0 to 9 9 represents 9 or more
57	III.3.C.	FLASHNOV	Canti-levered (or bridged) Flashing Lights-Not Over Traffic	1 N	266	0 to 9 9 represents 9 or more
58	III.3.D.	FLASHMAS	Mast Mounted Flashing Lights:	1 N	267	0 to 9 9 represents 9 or more
59	III.3.F.	FLASHOTH	Other Flashing Lights:	1 N	268	0 to 9 9 represents 9 or more

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER
(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
60	III.3.F.	FLASHDES	Specify:	9 C	269 (277)	Any Alphanumeric Description
61	III.3.G.	HWYSGNL	Hwy. Traffic. Signals	1 N	278	0 to 9 9 represents 9 or more
62	III.3.H.	WIGWAGS	Wigwags	1 N	279	0 to 9 9 represents 9 or more
63	III.3.J.	BELLS	Bells	1 N	280	0 to 9 9 represents 9 or more
64	III.4.	SPECPRO	Specify Warning Device:	20 C	281 (300)	Description of Non-train Activated Device
65	III.1.	NOSIGNS	No Signs or Signals	1 C	301	Enter a check if no signs or signals are present. 1=No signs or signals 0=At least one sign or signal
66	IV.10.	COMPOWER	Commercial Power Available (Y/N)?	1 C	302	1=Yes 2=No
67	III.7.	SGNLEQP	Signaling for Train Operation: Is Track Equipped with Train Signals	1 C	303	1=Yes 2=No
68	III.6.	SPSEL	Train Detection	1 C	304	New Values: 1= Constant Warning Time 3=DC/AFO 2= Motion Detectors 4=other 5=none (Previous Values: 1=Yes 2=No, 3=N/A) Conversion: Yes (1) CWT (1) No (2)-> DC/AFO(3) N/A (3)-> None (5) (Previous: Does Xing Signal Provide Speed Selection for Trains?)
69	IV.1.	DEVELTYP	Type of Development	1 C	305	Values are 1 to 5 1=Open Space 2=Residential 3=Commercial 4=Industrial 5=Institutional

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
70	IV.5	HWYPVED	Is Highway Paved?	1 C	306	1=Yes 2=No
71	IV.7.	DOWNST	Does Track Run Down a Street (Y/N)?	1 C	307	1=Yes 2=No
72	III.2.E.	PAVEMRK	Pavement Markings:	1 C	308	Values are 1 to 4 1=Stop lines, 2=RR Xing Symbols, 3=No Markings 4=Stop lines and RR Xing Symbols
73	IV.8.	HWYNEAR	Nearby Intersecting Highway?	1 C	309	New Values: 1=Less than 75ft 2=75 to 200ft 3=200 to 500 ft 4=N/A Previous Values: 1=Yes 2=No Conversion: Yes >Less than 75 ft. No >N/A (See Field 152, HWYNRSIG)
74	III.2.C.	ADVWARN	RR Advance Warning Signs	1 C	310	1=Yes 2=No
75	IV.2.	XANGLE	Smallest Crossing Angle	1 C	311	1 to 3 (measurement is in degrees) 1=0-29 2=30-59 3=60-90
76	IV.6.	SURFACE	Crossing Surface:	1 C	312	Conversion: New: 1. Timber; 2. Asphalt; 3. Asphalt & Flange; 4. Concrete; 5. Concrete and Rubber; 6. Rubber; 7. Metal; 8. Unconsolidated; 9. Other (Specify) Old: 1. Sectional Treated Timber; 2.Full Wood Plank; 3.Asphalt; 4.Concrete Slab; 5.Concrete Pavement; 6.Rubber; 7.Metal Sections; 8.Other Metal; 9.Unconsolidated; 0.Other (Specify) (See Field 151, XSUROTHR)
77	IV.3.	TRAFICLN	No. of Traffic Lanes Crossing RR:	1 C	313	Values are 1 to 9
78	IV.4.	TRUCKLN	Are Truck Pullout Lanes Present (Y/N)?	1 C	314	1=Yes 2=No

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER
(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
79	V.2.	STHWY1	Is crossing on State Highway System (Y/N)?	1 C	315	1=Yes 2=No
80	V.1.	HWYSYS	Highway System:	2 C	316 (317)	01=Interstate National Highway System 02=Other National Highway System 03=Other Federal-Aid Highway-Not NHS 08=Non Federal-Aid (NHS=National Highway System)
81	V.3.	HWYCLASS	Functional Classification of Road at Crossing:	2 C	318 (319)	01, 02, 06, 07, 08, 09, 11, 12, 14, 16, 17, 19 01=R. Interstate, 02=R. Oth. Prin. Arterial, 06=R. Minor Arterial, 07=R. Major Collector, 08=R. Minor Collector, 09=R. Local, 11=U. Interstate, 12=U. Oth. Freeway and Expressway, 14=U. Oth. Prin. Arterial, 16=U. Minor Arterial, 17=U. Collector, 19=U. Local [R=Rural, U=Urban]
82	V.5.	AADT	AADT	6 C	320 (325)	000001 – 999999 Annual Average Daily Traffic (AADT)
83	V.6.	PCTTRUK	Estimate Percent Trucks:	2 C	326 (327)	00 – 99 Estimate of % of Trucks
84	I.23.	LATITUDE	Latitude	10 7 N	328 (337)	Grade crossing latitudinal coordinate, from the center of the crossing.
85	I.24.	LONGITUD	Longitude	11 7 N	338 (348)	Grade crossing longitudinal coordinate, from the center of the crossing.
86	I.25.	LLSOURCE	Lat/Long Source	1 C	349	1 = actual 2=estimated Blank=neither 3. Federal Actual 4. Federal Derived –[For FRA Internal Use]
87	III.8.	INTRPRMP	Interconnection / Pre-emption	1 C	350	New values: 0 = not interconnected 1 = simultaneous preemption 2 = advance preemption 9 = n/a Previous values: 0 = not interconnected 1 = interconnected 2 = simultaneous preemption 3 = advance preemption 9 = n/a) Conversion: 1. (Interconnected)->1(simultaneous pre.) 2. (simulta. Pre.)->1(simultaneous pre.) 3. (adv.pre.)->2(adv pre.)
88	III.2.D.	HUMPSIGN	Hump Signs	1 C	351	1=Yes 2=No 3=Unknown Is Hump crossing sign is installed?
89	I.21.	HSCORRID	[High Speed] Corridor ID Code	2 C	352 (353)	Code must be in High Speed Corridor Table (obtain from FRA)
90		DOTACPD		8 6 N	354 (361)	DOT Accident Prediction Value

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER
(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
91		ACPDDATE		8 DATE	362 (369)	Indicates when DOT ACPD was generated.
92		ACCCNT1		2 N	370 (371)	Accident history – current complete year
93		ACCCNT2		2 N	372 (373)	Accident history – prior year
94		ACCCNT3		2 N	374 (375)	Accident history – two years prior
95		ACCCNT4		2 N	376 (377)	Accident history – three years prior
96		ACCCNT5		2 N	378 (379)	Accident history – four years prior
97		HISTDATE		8 DATE	380 (387)	Indicates when ACCCNT1- ACCCNT5 were generated
98	V.7.	SCHLBUS	Avg. No of School Buses Passing Over the Crossing on a School Day	3 N	388 (390)	Value must be 0 through 999
99	I.16	WHISTBAN	New: Whistle Ban (Quiet Zone)	1 C	391	Valid values: 0=no 1=24 hour 2=partial 9=unknown
100	I.19	PASSCD	Type of Passenger Service	1 C	392	Valid values: A = AMTRAK operates over crossing B = AMTRAK and other passenger train operates over crossing C = Other passenger train operates over crossing including Seasonal D = None
101	I.20	PASSCNT	Avg Passenger Train Count Per Day	3 N	393 (395)	Value must be 0 through 999. [Cannot exceed the total train movements]
102	I.10	RRMAIN	Parent RR	4 C	396 (399)	Valid Railroad Code
103	I.11	XINGOWNR	Crossing Owner	4 C	400 (403)	Valid Railroad or Company Code
104		SOURCE		1 C	404	This field will indicate the source of the last update. Valid values: H = other hard copy I = inventory form M = other magnetic media P = mass-update printout T = magnetic tape X = GX O = foreign files
105		UPDATDAT		8 DATE	405 (412)	This field will contain the date that the last update to the record was posted.

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION

INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)		DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
106		LONGBDAT		8 DATE	413	(420)	This field will contain the same date as the field EFFDATE, in this file, except that the year will be four characters in this data element.
107		LONGEDAT		8 DATE	421	(428)	This field will contain the same date as the field EDATE, in this file, except that the year will be four characters in this data element
108	III.3.B.	FOURQUAD	Four-quadrant gates present	1 C	429		1=Yes 2=No
109		TWOQUAD	Two-quadrant gates present	1 C	430		NOT USED IN NEW FORM
110	I.27.B.	OPENPUB	Private Crossing-Public Access	1 C	431		1=Yes 2=No Blank=Unknown
111	I.28.A.	RRNARR1	Railroad Use	20 C	432	(451)	These fields will contain whatever the railroad desires to enter.
112	I.28.B.	RRNARR2	Railroad Use	20 C	452	(471)	
113	I.28.C.	RRNARR3	Railroad Use	20 C	472	(491)	
114	I.28.D.	RRNARR4	Railroad Use	20 C	492	(511)	
115	I.29.A.	STNARR1	State Use	20 C	512	(531)	These fields will contain whatever the State desires to enter.
116	I.29.B.	STNARR2	State Use	20 C	532	(551)	
117	I.29.C.	STNARR3	State Use	20 C	552	(571)	
118	I.29.D.	STNARR4	State Use	20 C	572	(591)	
119	V.5	AADTYEAR	Year for AADT	4 C	592	(595)	This field will contain the year of the last AADT update.
120		AADTCALC		1 C	596		Not used.
121		TRAINDAT		4 C	597	600	Not currently used. Was to contain the year of the last trains update.
122		TRAINCAL		1 C	601		Not used. (This field was to identify how the last trains update was calculated: 1 = actual 2 = estimated Blank = neither)
123	III.9	RESERVE1	Reserved for Future Use	1 C	602		Reserved for future use. (RESERVE1 is 1 C. RESERVE2, RESERVE3, RESERVE4. and RESERVE5 are 3 C each.)

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)		DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
124	III.10	RESERVE2	Reserved for Future Use	3 C	603	(605)	
125	III.11	RESERVE3	Reserved for Future Use	3 C	606	(608)	
126	III.12	RESERVE4	Reserved for Future Use	3 C	609	(611)	
127	IV.11	RESERVE5	Reserved for Future Use	3 C	612	(614)	
128		DOTCASPD		8 6 N	615	(622)	DOT Predicted Casualty Rate
129		DOTFATPD		8 6 N	623	630)	DOT Predicted Fatality Rate
130		FUNCCAT		1 C	631		Not Used.
131	I.32.	RRCONT	Railroad Contact	10 C	632	(641)	This field contains the telephone number of the railroad contact associated with the crossing.
132	I.33.	HWYCONT	State Contact	10 C	642	(651)	This field contains the telephone number of the State highway contact associated with the crossing.
133	I.31.	POLCONT	Emergency Contact	10 C	652	(661)	This field contains the telephone number of the emergency contact associated with the crossing. Normally, this will be the ENS telephone number posted at the crossing or along the railroad branch line.
134	I.30.	NARR	Narrative	100 C	662	(761)	No editing will be done on this field
135	II.1.A.	TOTALTRN	Total Trains	3 N	762	(764)	0-500 Conversion: TOTALTRN = (DAYTHRU + DAYSWT + NGHTTHRU + NGHTSWT)
136	II.1.B.	TOTALSWT	Total Switching Trains	3 N	765	(767)	0-500 Conversion: TOTALSWT = DAYSWT + NGHTSWT
137	I.15.	ENSSIGN	ENS Sign	1 C	768		1 = Yes 2 = No
138	III.2.A.	XBUCK	Crossbucks	2 N	769	(770)	Conversion: XBUCK = XBUCKRF + XBUCKNRF
139	III.3.A.	GATES	Gates	2 N	771	(772)	Conversion: GATES = GATERW + GATEOTH
140	I.27.C.	PRVSIGNL	Signals -Specify	15 C	773	(787)	Conversion: If PRVIND = 2 then previous PRVSIGN value will be moved to PRVSGNL. (Refer to field 24 (PRVSIGN)

U.S. DOT CROSSING INVENTORY FORM

FORMAT FOR FRA INVENTORY FIELDS - DATA ENTRY FIELD DESCRIPTION
INVENTORY FIELD ORDER

(Fields not in Form 6180.71 are for FRA Internal Use)

FIELD/FORM 6180.71 NO.		FIELD NAME	DESCRIPTION	SIZE/ TYPE	START(END) (For ASCII)	DEFINITIONS, VALID VALUES, RANGES, & COMMENTS (CURRENT/NEW FIELDS ARE TO BE PROVIDED. PREVIOUS FIELDS ARE SHOWN IN THIS TABLE FOR INFORMATION ONLY.) {CONVERSIONS – FRA INTERNAL USE}
141	III.3.E.	FLASHPAI	Number of flashing light pairs	2 N	788 (789)	This field contains the number of flashing light pairs.
142	III.3.K.	WARNACTO	Other Train Activated Warning Devices	9 C	790 (798)	This field contains other train activated warning devices.
143	III.5.	CHANNEL	Channelization Devices with Gates	1 C	799	1=All Approaches 2=One Approach 3=None
144	I.26.	XINGADJ	Adjacent Xing with separate no.?	1 C	800	1=Yes 2=No
145	I.26.	XNGADJNO	Adjacent Xing with separate no.? Provide no.	7 C	801 (807)	Valid crossing number
146	IV.9.	ILLUMINA	Is Xing Illuminated?	1 C	808	1=Yes 2=No
147	V.4.	HWYSPEED	Posted Hwy Speed	3 N	809 (811)	This field contains the posted highway speed.
148		CNTYNAM	County	20 C	812 (831)	Valid County Name
149		TTSTNNAM	Nearest RR Timetable Station	25 C	832 (856)	Valid Timetable Station name
150		CITYNAM	City	20 C	857 (876)	Valid City Name
151	IV.6.	XSUROTHR	Crossing Surface: 9. Other	20 C	877 (896)	Specify Other Crossing Surface
152	IV.8.	HWYNRSIG	Nearby Intersecting Highway? Is it signalized?	1 C	897	1=Yes 2=No

Note: Data file submissions, must, at a minimum, contain the following data fields:

Initiating Agency (INIT),
Crossing Number (CROSSING),
Reason for Update (REASON),
Effective Date (EFFDATE),
State (FIPS Code) (STATE),

County (FIPS Code) (CNTYCD),
Railroad (RAILROAD),
Type of Crossing (TYPEXING),
Position of Crossing (POSXING),
and the updated data fields