Collision Risk Analysis of Freight Trains in the United States

(TRB 16-1250)

Accepted for Publication

Xiang Liu, Ph.D.
Assistant Professor
Department of Civil and Environmental Engineering
Rutgers, The State University of New Jersey
Phone: (848)-445-2868
Email: xiang.liu@rutgers.edu

ABSTRACT
Railroads support the national economy by carrying 43 percent of inter-city freight ton-miles in the United States. At the same time, train accidents damage infrastructure and rolling stock, disrupt operations, and may result in casualties and environmental damage. While the majority of previous studies focused on the safety risks associated with train derailments or highway-rail grade crossing incidents, much less work has been undertaken to evaluate train collision risk. This paper develops a statistical risk analysis methodology for freight-train collisions in the United States occurring between 2000 and 2014. Negative binomial regression models are developed to estimate the frequency of freight-train collisions as a function of year and traffic volume, both by track type and accident cause. Overall, the rate of train collision declined in the study period on both mainline and yard tracks. Train collision severity, as measured by the average number of railcars derailed, varied with the type of track and accident cause. Train collision risk, defined as the product of collision frequency and severity, is predicted for 2015 to 2017, based on the 2000 to 2014 safety trend. The statistical procedures developed in this paper can be adapted to various other types of consequences, such as damage costs or casualties. Ultimately, this paper and its sequent studies aim to provide the railroad industry with data analytic tools by which to discover useful information from historical accidents so as to make risk-informed safety decisions.

1 INTRODUCTION

The American economy hinges on freight railroads, which transport 43 percent of ton-miles of cargo. This reliance places importance on mitigating train accidents, which can damage infrastructure and rolling stock, disrupt operations, and cause casualties and environmental damage. There are three major accident types on U.S. freight-railroads: derailment, collision, and highway-rail grade crossing incidents. There is extensive research on derailment risk analysis (*1-11*) and highway-rail grade crossing safety (*12-14*). However, less work has been undertaken to evaluate train collision risk. To the authors' knowledge, few published studies are dedicated to statistical risk analysis of U.S. freight-train collisions. This knowledge gap motivates the development of this paper, which aims to address the following research inquires:

1) How can train collision risk be quantified?
2) How does train collision risk vary with time, traffic exposure, type of track, and accident cause?
3) How can collision risk be predicted?

2 DATA SOURCES

The Federal Railroad Administration (FRA) of the U.S. Department of Transportation (USDOT) requires railroads operating in the U.S. to submit detailed reports of accidents or incidents whose damage costs to track infrastructure, rolling stock, and signals exceeded a specified monetary threshold (*15*). The FRA compiles the submitted accident reports into a Rail Equipment Accident (REA) database. The REA database contains useful information on the type of railroad (e.g., freight railroad, passenger railroad), the type of accident or incident (e.g., derailment, collision, grade crossing incident, etc.), type of track (mainline, yard, siding, industrial), accident cause (e.g., track failures, mechanical failures, human errors), accident consequences (e.g., number of railcars derailed, track and rolling stock damage costs, casualties), and other accident circumstances. The FRA REA database has been used in many previous studies (*1-3, 11, 16*). In addition to reporting accident data, the railroads also report their monthly train-mile data through the FRA Operational Database. These data sources can be integrated and used to model freight-train collision risk. Note that each railroad carrier may also maintain an internal accident database containing those accidents that result in damage costs that fall below the FRA mandatory reporting threshold. The Non-FRA-reportable accidents are typically not publicly available, and thus are excluded from the analysis.

3 RESEARCH OBJECTIVES AND SCOPE

This paper aspires to make the best use of historical FRA-reportable freight-train collision data to understand major collision causes and the temporal change in train collision risk. The risk analysis will ultimately inform decision-makers in the process of data-driven safety policy development. Specifically, this research aims to produce the following deliverables:

- Develop a quantitative methodology for freight-train collision risk analysis
- Develop statistical procedures and toolboxes for modeling the frequency and severity of freight-train collision by influencing factors
- Project future train collision risk based on the current safety trend. The risk projection provides a scientific basis for evaluating the safety benefit of

prospective collision avoidance technologies

The majority of train accidents occur on Class I railroads (*5*). Each Class I railroad has annual operating revenue exceeding $478.5 million (2014 dollars). Class I railroads accounted for approximately 68% of U.S. railroad route miles, 97% of total ton-miles transported, and 94% of the total freight rail revenue (*17*). Therefore, this paper focuses on Class I railroad freight-train collisions from 2000 to 2014. Depending on data availability and questions of interest, the methodologies can be adapted to other types of train accident.

The analyses in this paper focus on collisions between freight trains, as well as on collisions between freight trains and non-train consists, such as maintenance vehicles, cut of cars, and locomotives. However, this paper excludes collisions between freight trains and passenger trains. This type of collision needs to account for both freight and passenger train operations and their possible interactions. Modeling this type of freight-and-passenger train collision requires a separate research study.

This paper is structured as follows. First, some key definitions are presented to clarify the scope of this analysis. Second, statistical models are developed to estimate collision frequency as a function of year and traffic volume, both by track type and accident cause. Third, collision severity (measured by the average number of railcars derailed per train collision) is estimate based on historical accident data. Fourth, future train collision frequency and severity are projected based on the 2000 to 2014 safety trend. Finally, the paper discusses the implications of the analyses with respect to research and practice.

4 DEFINITIONS

Collision. According to the FRA, a collision is defined as "an impact between on-track equipment consists while both are on rails and where one of the consists is operating under train movement rules or is subject to the protection afforded to trains." (*15*) This definition includes instances in which a portion of a consist occupying a siding fouls the mainline and is struck by an approaching train. In the FRA REA database, there are six types of collision: head-on, rear-end, side, raking, broken train, and railroad crossing. A head-on collision occurs when the trains or locomotives involved are traveling in opposite directions on the same track. A rear-end collision occurs when the trains or locomotives involved are traveling in the same direction on the same track. A side collision occurs at a turnout where one consist strikes the side of another consist. A raking collision occurs between the parts or lading of a consist on an adjacent track, or with a structure such as a bridge. A broken train collision occurs when a moving train breaks into parts and an impact occurs either between these parts or between a portion of the broken train and another consist. A railroad crossing collision occurs between on-track railroad equipment at a point where tracks intersect (*15*).

Safety. Safety can be defined as the number of accidents, evaluated by kind and severity, which are expected to occur on the entity during a specified period (*18*). Liu (*11*) provides a theoretical explanation for this definition from a stochastic process perspective. One highlight of the safety definition is "expected to occur." The difference between the observed and expected number of accidents represents the statistical uncertainty of accident occurrence (*11*). The expected rate of accident frequency or accident severity can be estimated using multivariate regression techniques, which will be detailed later.

Risk. The risk can be defined as the likelihood and consequence of an incident (*19*). In practice, researchers sometimes use the expected consequence to measure the risk (3, 6, 7). For example, if the accident consequence is measured by damage cost, the risk is interpreted by the expected damage cost.

The above-mentioned definitions form the basis of the following statistical analyses. Depending on data availability and questions of interest, analysts may choose to use alternative definitions. If so, they can adapt this paper's methodology accordingly.

5 STATISTICAL METHODOLOGY FOR TRAIN COLLISION RISK ANALYSIS

5.1 Model Development

The FRA REA database records four types of track: main, siding, yard, and industrial tracks. These track types are used for different operational functions and consequently entail different accident types, causes, and consequences (*5*). Train accidents are categorized according to derailment, collision, highway–rail grade crossing incident, and several other less frequent types. Liu et al. (*5*) presents an analysis of train derailment frequency and severity by the type of track and type of accident using data from 2000 to 2010. Table 1 shows an updated analysis that includes more recent data. In this paper, the collisions include those between freight trains, as well as those between freight trains and non-train consists (e.g., maintenance vehicles, cut of cars, locomotives).

The analysis shows that collisions accounted for approximately 6 percent of total accident frequency and 4 percent of total number of railcars derailed on Class I freight railroads. Derailment was the most common type of accident on each track type, accounting for 72 percent of train accidents and 94 percent of the total number of railcars derailed, across all types of track. On main tracks, the frequency of freight-train collisions is 80 percent lower than in grade crossing incidents (429 versus 1,929). However, on mainlines the average number of railcars derailed per collision is nine times greater than in grade crossing incidents (4.3 versus 0.5). Ninety-nine percent of highway–rail grade crossing incidents occurred on main tracks and accounted for 21% of all types of accidents on the main tracks of Class I freight railroads. Chadwick et al. (*14*) found that many grade crossing incidents exceeded the FRA reporting threshold for monetary damages, but did not result in a derailment.

TABLE 1 Accident Frequency and Severity by Accident Type and Track Type, U.S. Freight Railroads, 2000–2014

Number of Freight Train Accidents

	Derailment	Collision	Highway-Rail	Other	Total
Main	6,026	429	1,929	874	9,258
Yard	4,220	524	14	518	5,276
Siding	632	33	7	66	738
Industry	1,286	76	9	190	1,561
Total	12,164	1,062	1,959	1,648	16,833

Number of Railcars Derailed in Freight Train Accidents

	Derailment	Collision	Highway-Rail	Other	Total
Main	51,993	1,793	901	685	55,372
Yard	19,763	998	10	737	21,508
Siding	3,353	116	5	68	3,542
Industry	5,793	121	12	119	6,045
Total	80,902	3,028	928	1,609	86,467

Average Number of Railcars Derailed per Train Accident

	Derailment	Collision	Highway-Rail	Other	Total
Main	8.6	4.2	0.5	0.8	6.0
Yard	4.7	1.9	0.7	1.4	4.1
Siding	5.3	3.5	0.7	1.0	4.8
Industry	4.5	1.6	1.3	0.6	3.9
Total	6.7	2.9	0.5	1.0	5.1

Note: The number of derailed railcars includes both empty and loaded railcars. Where there are multiple trains involved in a collision, the total number of railcars derailed from all the trains is counted as the collision severity.

A statistical model is developed to correlate train collision frequency with influencing factors. Based on data from the FRA REA database, this paper focuses on two potential affecting factors: year and annual traffic exposure. The “year” variable tests whether there is a temporal change in train collision frequency given traffic exposure. The “traffic exposure” variable describes whether and how train collision count varies with traffic volume in a given year. A basic model structure is as follows:

$$\mu_i = \exp(\alpha+\beta\times T_i+ \theta\times M_i)M_i \quad (1)$$

where:

μ_i = expected number of freight-train collisions in year i

T_i = year (for example, T_i is 2000 for year 2000)

M_i = million train-miles in year i
α, β, θ = parameter coefficients

Collision rate is defined as the number of train collisions, normalized by traffic exposure. According to this definition, Equation (1) is re-written as:

$$Z_i = \mu_i / M_i = \exp(\alpha + \beta \times T_i + \theta \times M_i) \tag{2}$$

A similar model was used in several previous studies (*11, 20-22*). Those studies assumed that train accident rate is independent of traffic exposure. In order to understand whether and how train collision frequency varies with traffic volume, our model generalizes the previous model by introducing a new parameter θ. $\theta > 0$ means that if traffic increases, collision rate will increase with traffic volume, given all else being equal. The previous model (*11, 20-22*) is a special case of the generalized model, given $\theta = 0$.

The next step is to estimate the unknown parameters based on historical data. The literature contains numerous regression models, of which the negative binomial regression is used frequently. A technical review of the negative binomial model can be found in (*23*). This paper starts with negative binomial regression. Where the goodness of fit is inadequate, alternative models will be used. The negative binomial regression results are presented in Table 2. The last column is the P-value of a parameter estimator, which represents the statistical significance of a predictor variable using the Wald Test (*23*). A generally acceptable rule is that if a predictor variable has a P-value smaller than 5 percent, this variable is significant. The analysis found that the parameter coefficient for the variable "year" is significantly negative ($\beta = -0.0414$; $P < 0.0001$), indicating that there is a significant temporal decline in train collision rate, given traffic exposure.

TABLE 2 Negative Binomial Regression of Train Collision Frequency, Class I Freight Railroads, All Types of Track, 2000 to 2014

Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
α	79.6894	17.3774	45.6303	113.7485	21.03	<.0001
β	-0.0414	0.0086	-0.0582	-0.0246	23.25	<.0001
θ	0.0020	0.0010	0.0001	0.0039	4.07	0.0438

Deviance = 13.3, Degree of Freedom = 12, P = 0.35

The goodness-of-fit of a negative binomial model can be evaluated using a statistical criterion called "Deviance." Statistical theory dictates that the Deviance asymptotically follows a Chi Square distribution (*23*). Based on this property, the P-value in the deviance test can be calculated. In general, if the P-value in the deviance test is larger than 5 percent, the model will appear to be an adequate fit to the empirical data. In our example, the deviance is 13.3, the degree of freedom is 12 and the corresponding P-value = 0.35 (Table 2). Based on the fitted parameters, the expected frequency of U.S. freight-train collision is estimated using the following equation:

$$\mu_i = \exp(79.6894-0.0414T_i+0.0020M_i)M_i \tag{3}$$

Equation (3) is written in a mathematically equivalent way:

$$Z_i = Y_i / M_i = \exp(79.6894-0.0414T_i+0.0020M_i) \tag{4}$$

Equation (4) estimates Class I railroad freight-train collision rate on all types of track. The analysis shows that the expected annual collision rate declines between 2000 and 2014. In addition to the temporal change, collision rate increases with traffic volume given all else being equal. Table 3 compares the observed versus the predicted number of freight-train collisions using the negative binomial regression model described above.

TABLE 3 Empirical versus Predicted Freight-Train Collision Frequency, All Tracks, Class I Freight Railroads, 2000 to 2014

Year	Total Train-Miles (Million)	Observed Collision Frequency	Estimated Collision Frequency
2000	599.97	91	92
2001	580.90	97	82
2002	596.68	65	83
2003	610.74	76	84
2004	634.08	101	88
2005	649.45	106	89
2006	666.29	74	91
2007	642.99	82	80
2008	621.41	65	71
2009	516.87	46	46
2010	554.00	45	51
2011	566.12	52	51
2012	577.27	56	51
2013	589.85	52	51
2014	605.01	54	52

In addition to the Deviance, another common goodness-of-fit test is the Chi-square test, which assesses the relative difference between each observation and estimation.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{5}$$

Where:

O_i = observed number of collisions in year i
E_i = estimated number of collisions in year i
n = sample size (number of years)

Based on Table 3, χ^2 = 17.7. The corresponding P-value is 0.22 (degree of freedom is 14). Therefore, it indicates that the estimated collision frequency reasonably matches the observed count. Both the Deviance test and the Chi-square test show that a negative binomial regression model can be used to fit the empirical freight-train collision data in this paper. In the next section, this regression technique is applied to major collision causes on mainline and yard tracks.

5.2 Major Collision Causes

The FRA REA database records over 300 accident cause codes. Each cause code describes a specific accident circumstance. The train accident cause codes are hierarchically organized and categorized into major cause groups: track, equipment, human factors, signal, and miscellaneous (*15*). Within each cause group, FRA organizes individual cause codes into subgroups of related causes, such as roadbed and track geometry, within the track group and similar subgroups within the other major cause groups. A variation on the FRA subgroups was developed by Arthur D. Little (ADL), in which similar cause codes were combined into groups on the basis of expert opinion (*24*). The ADL groupings are similar to FRA's subgroups but are more fine-grained for certain causes, thereby allowing greater resolution in some cases. For example, the FRA grouping combines broken rails, joint bars, and rail anchors in the same subgroup, whereas the Arthur D. Little grouping distinguishes between broken rail and joint bar defects. These groups were used to analyze cause-specific collision frequency and severity. Note that the ADL accident cause grouping might not be the only grouping approach. Furthermore, the same cause may fall into multiple groups. Therefore, if analysts use a different accident cause grouping scheme, the analyses should be adapted accordingly.

According to Figure 1, on main tracks (including siding tracks thereafter), failure to obey or display signals and violation of train speed rules are the top two collision causes, whereas on yard tracks (including industrial tracks thereafter),violation of switching rules is the top cause group. A detailed breakdown of the cause codes within each ADL cause group can be found in ADL (*24*).

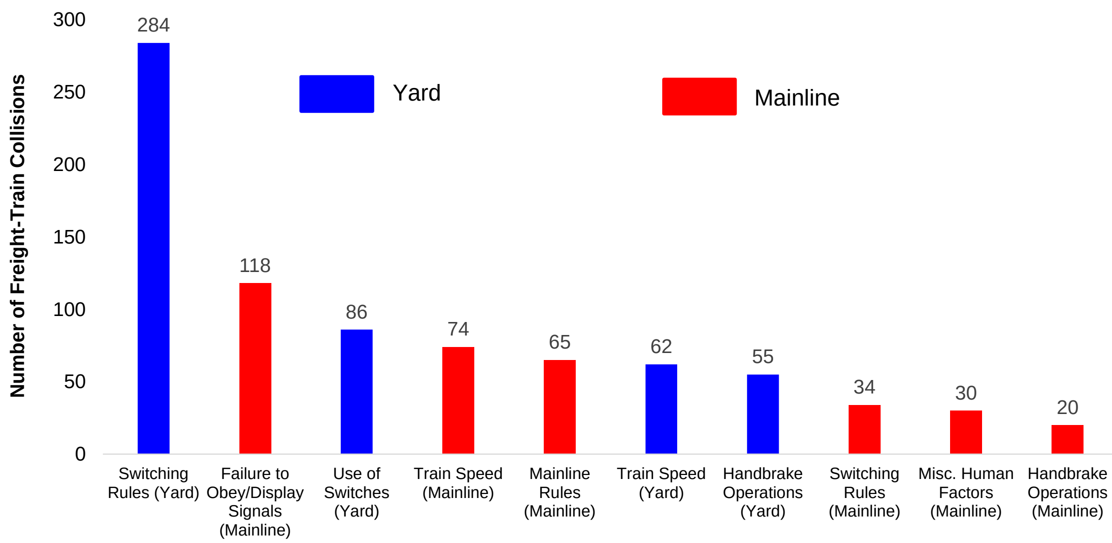

FIGURE 1 Top collision cause group (ADL cause group) on mainline and yard tracks, Class I freight railroads, 2000 to 2014.

Using the negative binomial regression described above, a cause-specific train collision frequency model is developed (Table 4). All the models presented in Table 4 have been validated based on the Deviance test and Chi-square test. Each model represents the "best" fit to the empirical data.

TABLE 4 Selected Freight-Train Collision Frequency Regression Models

	Collision Frequency (μ_i) by Year (T_i) and Million Train-Miles (M_i)
Failure to obey/display signals on main tracks	$\mu_i = \exp(101.27 - 0.0526T_i)M_i$
Violation of train speed rules on main tracks	$\mu_i = \exp(-9.90 + 0.0096M_i)M_i$
Total mainline collisions	$\mu_i = \exp(83.61 - 0.0442T_i + 0.0039M_i)M_i$
Violation of switching rules on yard tracks	An average of 19 collisions per year, no significant temporal and traffic effects
Total yard collisions	$\mu_i = \exp(90.28 - 0.0453T_i)M_i$
Total train collisions on all types of tracks	$\mu_i = \exp(79.69 - 0.0414T_i + 0.002M_i)M_i$

6 COLLISION SEVERITY

In addition to collision frequency, severity is another important element in train collision risk analysis. This paper uses the average number of railcars derailed per freight-train collision as a proxy to measure collision severity. This metric is related to accident kinetic energy and has been extensively used in the prior work (*1*, *4*-*8*). On average, mainline collisions tend to derail more railcars as compared to collisions on yard and industrial tracks. A Wald-Wolfowitz runs test (*25*) was used to understand whether there is any significant temporal trend in collision severity. This statistical test checks if a data set results from a random process. When the P-value in the test is larger than 0.05, we may conclude that there is no statistically significant trend in terms of the average annual collision severity. Table 5 shows that the severities for all of the selected cause groups have no significant temporal trend. The yearly severity fluctuation is largely due to random variations. Therefore, in the following risk analysis, we will use the average of collision severity by type of track and accident cause (Table 5).

TABLE 5 Number of Railcars Derailed per Train Collision, Class I Freight Railroads, 2000 to 2014

		Mainline			Yard	
Year	All Tracks, All Causes	Failure to Obey/Display Signals	Violation of Train Speed Rules	All Causes on Mainlines	Violation of Switching Rules	All Causes on Yard Tracks
2000	1.9	1.9	2.4	1.7	1.3	2.1
2001	3.9	19.0	2.8	7.2	1.0	1.5
2002	3.6	12.8	1.7	7.5	0.4	1.0
2003	2.5	8.0	8.3	3.9	1.3	1.4
2004	3.7	7.9	4.5	5.5	1.6	2.1
2005	3.2	5.6	4.8	4.4	0.9	1.7
2006	2.1	5.6	4.0	3.0	1.6	1.4
2007	2.8	7.2	0.7	4.6	1.4	1.5
2008	2.7	10.9	3.0	4.8	1.1	1.5
2009	1.7	5.2	0.7	2.6	1.0	1.2
2010	1.9	3.8	1.7	2.7	1.3	1.4
2011	2.3	5.4	0.5	2.9	1.0	1.9
2012	2.5	7.8	23.0	4.0	0.7	0.9
2013	3.8	17.8	0.0	6.9	1.4	1.3
2014	3.0	2.5	6.0	6.0	0.9	1.3
Average	2.8	8.1	4.3	4.5	1.1	1.5
Standard Error	0.2	1.3	1.5	0.5	0.1	0.1
P-value in Runs Test	0.11	0.09	0.16	0.25	0.20	0.20

7 COLLISION RISK ANALYSIS

Train collision risk can be defined as the product of collision frequency and severity:

$$R = F \times S \tag{6}$$

Where:
R = estimated annual collision risk
F = estimated annual collision frequency
S = estimated collision severity (number of railcars derailed per train collision)

Both the estimated frequency and severity are subject to statistical uncertainty. Correspondingly, there is uncertainty associated with the risk estimator. The variance of the risk estimator can be calculated using the following equation by assuming that the variances of the estimators of collision frequency and severity are independent. A statistical proof of the variance of two random variables can be found in Goodman (*26*),

$$Var(R) = Var(F \times S) = Var(F)Var(S) + Var(F)E(S)^2 + Var(S)E(F)^2 \tag{7}$$

Where:
Var(R) = variance of collision risk estimator
Var(F) = variance of collision frequency estimator
Var(S) = variance of collision severity estimator
E(F) = the estimator of collision frequency
E(S) = the estimator of collision severity

Furthermore, the 95-percent confidence interval of the collision risk estimator ($CI_{95\%}(R)$) is:

$$CI_{95\%}(R) = [R - 1.96\sqrt{Var(R)}, R + 1.96\sqrt{Var(R)}] \tag{8}$$

Based on Equations (6) to (8), the estimated annual collision risk and its 95-percent confidence interval is calculated (Table 6) using the average traffic volume between 2000 and 2014. For example, if the traffic volume in 2017 is 533 million train-miles on Class I mainlines, the projected collision frequency is 19, with a 95-percent confidence interval of 14 to 25. This projection entails a 95-percent chance that the actual number of collisions will be between 14 and 25. Collision risk, measured by the expected number of railcars derailed, is estimated to range between 55 and 116 in year 2017. At the time of analysis, the future traffic volume was not available. A sensitivity analysis was conducted to predict the range of train collision risk at different traffic levels (Table 6). It shows that collision risk varies with traffic volume. The change of mainline risk is more sensitive to traffic than the risk on yard track. For example, if the traffic volume increases from 533 to 549 million train-miles on mainlines in year 2015 (3 percent increase), the estimated collision risk increases from 95 to 104, at a 9-percent increase (scenarios 2 and 3 in Table 6). By contrast, the collision risk on yard track increases at 3 percent. This is because mainline collision frequency increases exponentially with traffic volume in a given year, whereas yard collision frequency increases linearly with traffic, according to the regression models in Table 4.

Note that implementation of certain collision avoidance technologies (e.g., positive train control) may change the safety trend described in this paper. We recommend that all train accident analyses be periodically updated to reflect up-to-date safety statuses. As in the highway safety community, the change of risk before and after the installation of a specific safety measure could be used to evaluate that particular measure's safety benefit. A before-after safety evaluation of train collision risk mitigation strategies is the next step of this research.

TABLE 6 Projected Collision Frequency, Severity and Risk in 2015 to 2017, Class I Freight Railroads, at Different Traffic Volumes

Scenario 1: Baseline traffic decreases by 3 percent

(a) Mainline		Collision Frequency			Collision Severity (Number of Cars Derailed per Collision)			Collision Risk (Total Numer of Cars Derailed)		
Year	Mainline Million Freight-Train Miles	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound
2015	517.0	19	15	24	4.5	3.6	5.4	86	59	112
2016	517.0	18	14	23	4.5	3.6	5.4	81	55	107
2017	517.0	17	13	23	4.5	3.6	5.4	77	49	104

(b) Yard		Collision Frequency			Collision Severity (Number of Cars Derailed per Collision)			Collision Risk (Total Numer of Cars Derailed)		
Year	Yard Switching Million Train-Miles	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound
2015	65.5	27	22	32	1.5	1.3	1.7	41	31	50
2016	65.5	26	21	31	1.5	1.3	1.7	39	30	48
2017	65.5	24	20	30	1.5	1.3	1.7	36	27	45

Scenario 2: Baseline traffic volume (average traffic volume between 2000 and 2014)

(a) Mainline		Collision Frequency			Collision Severity (Number of Cars Derailed per Collision)			Collision Risk (Total Numer of Cars Derailed)		
Year	Mainline Million Freight-Train Miles	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound
2015	532.99	21	16	26	4.5	3.6	5.4	95	65	124
2016	532.99	20	15	26	4.5	3.6	5.4	90	59	121
2017	532.99	19	14	25	4.5	3.6	5.4	86	55	116

(b) Yard		Collision Frequency			Collision Severity (Number of Cars Derailed per Collision)			Collision Risk (Total Numer of Cars Derailed)		
Year	Yard Switching Million Train-Miles	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound
2015	67.53	27	23	33	1.5	1.3	1.7	41	31	50
2016	67.53	26	21	32	1.5	1.3	1.7	39	29	49
2017	67.53	25	20	31	1.5	1.3	1.7	38	28	47

<u>Scenario 3: Baseline traffic volume increases by 3 percent</u>

(a) Mainline		Collision Frequency			Collision Severity (Number of Cars Derailed per Collision)			Collision Risk (Total Numer of Cars Derailed)		
Year	Mainline Million Freight-Train Miles	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound
2015	548.98	23	17	29	4.5	3.6	5.4	104	69	138
2016	548.98	22	16	29	4.5	3.6	5.4	99	64	134
2017	548.98	21	15	28	4.5	3.6	5.4	95	60	129

(b) Yard		Collision Frequency			Collision Severity (Number of Cars Derailed per Collision)			Collision Risk (Total Numer of Cars Derailed)		
Year	Yard Switching Million Train-Miles	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound	Mean	95% Lower Bound	95% Upper Bound
2015	69.56	28	23	34	1.5	1.3	1.7	42	32	52
2016	69.56	27	22	33	1.5	1.3	1.7	41	31	50
2017	69.56	26	21	32	1.5	1.3	1.7	39	29	49

8 DISCUSSIONS

In this section, we discuss the contributions of this study with respect to the literature and practice. We also discuss the data limitations of the current research and suggest possible future research directions.

8.1 Contributions to the literature

Because the FRA has been collecting train accident data since the 1970s, researchers are able to look into historical accident data, discover useful information, and propose risk-informed decisions. Compared with the highway safety community, in which statistical modeling of accident data is normative for research and policy making, there has been much less statistical modeling work in the U.S. railroad sector. While most existing railroad safety studies have concentrated on derailments or grade crossing incidents, very limited statistical research has been spent on train collision risk analysis. This paper intends to develop an implementable statistical methodology for estimating freight-train collision frequency and severity. One of the most important lessons learned by transportation safety analysts in the past decades has been a statistical phenomenon called "Regression to the Mean" (RTM). The RTM refers to the tendency of a random variable that deviates from the mean to return to "normal" given nothing has changed (*18*). In the context of rail safety, RTM implies that a high accident rate in one year may be followed by a low rate in the next year due to the random fluctuation, even if there is no actual safety change (*11*). As Liu (*11*) mathematically proves, the RTM is inherent in any empirical train accident data and must be addressed through statistical approaches in order to understand the "true" safety trend. The collision analysis presented in this paper provides a step-by-step procedure for identifying the data-driven safety performance function (SPF) in the railroad industry, accounting for random fluctuations in accident occurrence and severity. The general approaches and procedures herein can be adapted to other types of train accidents and severities.

8.2 Contributions to the practice

The Rail Safety Improvement Act (RSIA) requires railroads to adopt risk-based approaches to ensure operational safety. However, there is no definitive, normative, practical methodological framework to guide the process of risk assessment. By contrast, the Federal Highway Safety Administration (FHWA) publishes an extensive manual to guide the statistical modeling of vehicle crash data, based on decades of research (*27*). The highway safety manual (HSM) provides practitioners with easy-to-understand tutorials on the basic concepts of transportation statistics and with toolboxes for statistically analyzing and interpreting accident data. The author hopes to use this paper, in conjunction with its sequent studies, to promote industry-academic-wide dialogues in order to develop a railroad safety analysis manual, just as the highway sector has done over the past decade. One primary application of the statistical modeling of safety data is to evaluate potential collision avoidance technologies. By calculating the accident prediction models before and after implementation of certain risk reduction strategies, decision makers can better understand how a specific technology changes the safety trend, as well as the magnitude of its safety benefit. Ultimately, the railroad research community could develop a data-driven guideline for optimal safety investment.

9 CONCLUSIONS

This paper develops a statistical risk analysis of freight-train collisions in the United States, based on data from 2000 to 2014. The analysis shows that there is a temporal decline in collision rate on both mainline and yard tracks during the study period. The relationship between annual collision frequency and traffic exposure may vary with the type of track and accident cause. The statistical model can be used to project freight-train collision risk in the future, enabling a data-driven assessment of the safety effectiveness of certain accident prevention strategies.

10 NEXT STPES

First, the methodology will be adapted to account for other collision consequences, such as casualties, property damages, or environmental impact. These modifications will enable a comprehensive risk evaluation of train collision risk. Second, the methodology can be applied to passenger or rail transit accident analyses. Third, there is recent concern regarding crude oil transportation risk. The collision frequency and severity prediction models will be incorporated into a hazardous materials risk analysis model to estimate the amount of affected population due to a potential crude oil release. Fourth, a before-versus-after safety evaluation framework can be developed to assess how certain technologies affect collision risk. Finally, this paper focuses on freight-train collision accident statistics in the United States. Other researchers studied train accidents in Canada (*3*, *28*, *29*) or Europe (*22*). Although the full comparison between U.S. rail safety statistics and those in other regions is beyond the main scope of this paper, it might be interesting to adapt the statistical methodology to other rail systems in future research.

Acknowledgements

The author was partially funded by the Department of Civil and Environmental Engineering (CEE), the Center for Advanced Infrastructure and Transportation (CAIT), and the School of Engineering (SOE), all at Rutgers University. The author thanks Shuang Xu and Yanlei Wang for their assistance. However, the author is solely responsible for all the views and analyses herein.

REFERENCES CITED

1. Barkan, C.P.L., Dick, C.T., Anderson, R. Railroad derailment factors affecting hazardous materials transportation risk. *Transportation Research Record 1825*, 2003, pp. 64–74.
2. Anderson, R.T., Barkan, C.P.L. Railroad accident rates for use in transportation risk analysis. *Transportation Research Record 1863*, 2004, pp. 88–98.
3. Bagheri, M., Saccomanno, F.F., Chenouri, S., Fu, L.P. Reducing the threat of in transit derailments involving dangerous goods through effective placement along the train consist. *Accident Analysis and Prevention*, Vol. 43, No. 3, 2011, pp. 613–620.
4. Liu, X., Barkan, C.P.L., Saat, M.R. Analysis of derailments by accident cause: evaluating railroad track upgrades to reduce transportation risk. *Transportation Research Record 2261*, 2011, pp. 178–185.
5. Liu, X., Saat, M.R., Barkan, C.P.L. Analysis of causes of major train derailment and their effect on accident rates. *Transportation Research Record 2289*, 2012, pp. 154–163.
6. Liu, X., Saat, M.R., Barkan, C.P.L. Integrated risk reduction framework to improve railway hazardous materials transportation safety. *Journal of Hazardous Materials*, Vol. 260, 2013a, pp. 131–140.
7. Liu, X., Saat, M.R., Barkan, C.P.L. Safety effectiveness of integrated risk reduction strategies for rail transport of hazardous materials. *Transportation Research Record 2374*, 2013b, pp. 102–110.
8. Liu, X., Saat, M.R., Qin, X., Barkan, C.P.L. Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Analysis and Prevention*, Vol. 59, 2013c, pp. 87–93.
9. Liu, X., Saat, M.R., Barkan, C.P.L. Probability analysis of multiple-tank-car release incidents in railway hazardous materials transportation. *Journal of Hazardous Materials*, Vol. 276, 2014a, pp. 442–451.
10. Liu, X., Lovett, A., Dick, C.T., Saat, M.R., Barkan, C.P.L. Optimization of rail defect inspection frequency for the improvement of railway transportation safety and efficiency. *Journal of Transportation Engineering*, Vol. 140, No. 10, 2014b, 04014048.
11. Liu, X. Statistical temporal analysis of freight-train derailment rates in the United States: 2000 to 2012. *Transportation Research Record 2476*, 2015, pp. 119-125.
12. Austin, R.D., Carson, J.L. An alternative accident prediction model for highway-rail interfaces. *Accident Analysis and Prevention*, Vol. 34, 2002, pp. 31–42.
13. Saccomanno, F.F., Park, P.Y.-J., Fu, L. Estimating countermeasure effects for reducing collisions at highway-railway grade crossings. *Accident Analysis and Prevention*, Vol. 39, 2007, pp. 406–416.
14. Chadwick, S.G., Saat, M.R., Barkan, C.P.L. Analysis of factors affecting train derailments at highway-rail grade crossings. In: Proceedings of the Transportation Research Board 91st Annual Meeting, 2012.
15. Federal Railroad Administration (FRA). *Guideline for accident report preparation*. FRA, Washington D.C., 2011.
16. Nayak, P.R., Rosenfield, D.B., Hagopian, J.H. *Event probabilities and impact zones for hazardous materials accidents on railroads.* Report DOT/FRA/ORD-83/20, FRA, U.S. Department of Transportation, Washington, D.C., 1983.

17. Association of American Railroads (AAR). *Class I railroad statistics*, 2013.
18. Hauer, E. *Observational Before-After Studies in Road Safety*. Pergamon Press, Elsevier Science Ltd., Oxford, England, 1997.
19. Kaplan, S., Garrick, B.J. On the quantitative definition of risk. *Risk Analysis*, Vol. 1, No. 1, 1981, pp. 11-27.
20. Evans, A.W. Rail Safety and Rail Privatization in Great Britain. *Accident Analysis and Prevention*, Vol. 39, No. 3, 2007, pp. 510-523.
21. Evans, A.W. Rail Safety and Rail Privatization in Japan. *Accident Analysis and Prevention*, Vol. 42, No. 4, 2010, pp. 1296-1310.
22. Evans, A.W. Fatal Train Accidents on Europe's Railways: 1980-2009. *Accident Analysis and Prevention*, Vol. 43, No. 1, 2011, pp. 391-401.
23. Hilbe, J.M. *Negative binomial regression*, Cambridge University Press, 2007.
24. Arthur D. Little, Inc. (ADL). *Risk Assessment for the Transportation of Hazardous Materials by Rail. Supplementary Report: Railroad Accident Rate and Risk Reduction Option Effectiveness Analysis and Data*, second revision. Cambridge, Mass, 1996.
25. Bradley. *Distribution-Free Statistical Tests*. Prentice-Hall, 1968.
26. Goodman, L.A. The variance of the product of K random variables. *Journal of the American Statistical Association*, 1962.
27. American Association of State Highway and Transportation Officials (AASHTO). *An introduction to highway safety manual*, 2015. http://www.highwaysafetymanual.org/Documents/HSMP-1.pdf.
28. Saccomanno, F.F., El-Hage, S.M. Minimizing derailments of railcars carrying dangerous commodities through effective marshaling strategies. *Transportation Research Record 1245*, 1989, pp. 34–51.
29. Saccomanno, F.F., El-Hage, S.M. Establishing derailment profile by position for corridor shipments of dangerous goods. Canadian Journal of Civil Engineering, Vol. 18, No. 1, 1991, pp. 67–75.