Analysis of Collision Risk for Freight Trains in the United States

Xiang Liu

Railroads support the national economy by carrying 43% of intercity 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 methodology for risk analysis of freight train collisions in the United States occurring between 2000 and 2014. Negative binomial regression models were 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 collisions declined in the study period on both main line and yard tracks. Severity of train collisions, 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 on the basis of the safety trend in 2000 to 2014. The statistical procedures developed in this paper can be adapted to various other types of consequences, such as damage costs and casualties. Ultimately, this paper and its sequent studies aim to provide the railroad industry with data analytics tools by which to discover useful information from historical accidents so as to make risk-informed safety decisions.

The American economy hinges on freight railroads, which transport 43% of annual 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. Three major accident types occur on U.S. freight railroads: derailment, collision, and highway–rail grade-crossing incidents. The research on derailment risk analysis (*1–11*) and highway–rail grade crossing safety (*12–14*) is extensive. However, less work has been undertaken to evaluate train collision risk. To the author's knowledge, few published studies are dedicated to statistical analysis of risk for U.S. freight train collisions. This knowledge gap motivated development of this paper, which aims to address the following research inquiries:

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?

Department of Civil and Environmental Engineering, School of Engineering, Rutgers University, 96 Frelinghuysen Road, Room 607, Piscataway, NJ 08854. xiang.liu@rutgers.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2546, Transportation Research Board, Washington, D.C., 2016, pp. 121–128.
DOI: 10.3141/2546-15

DATA SOURCES

FRA of the U.S. Department of Transportation requires railroads operating in the United States to submit detailed reports of accidents or incidents whose damage costs to track infrastructure, rolling stock, and signals exceeded a specified monetary threshold (*15*). FRA compiles the submitted accident reports into a database of rail equipment accidents (REAs). The REA database contains useful information on the type of railroad (e.g., freight, passenger), the type of accident or incident (e.g., derailment, collision, grade crossing incident), type of track (main line, 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 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 database of operations. These data sources can be integrated and used to model freight train collision risk. Each railroad carrier may also maintain an internal accident database containing those accidents that result in damage costs below the FRA mandatory reporting threshold. The non-FRA-reportable accidents are typically not publicly available, and thus are excluded from the analysis.

RESEARCH OBJECTIVES AND SCOPE

This paper aspired 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 developing data-driven safety policies. Specifically, this research aimed to produce the following deliverables:

- A quantitative methodology for analyzing the risk of freight train collisions,
- Statistical procedures and toolboxes for modeling the frequency and severity of freight train collisions by influencing factors, and
- Projection of the future risk of train collisions on the basis of current safety trends. 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.

As the methodologies depend on data availability and questions of interest, they can be adapted to other types of train accidents.

The analyses in this paper focus on collisions between freight trains as well as on those between freight trains and nontrain consists, such as maintenance vehicles, cut of cars, and locomotives. However, this paper excludes collisions between freight trains and passenger trains because analysis of such collisions must 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 the analysis. Second, statistical models are developed to estimate collision frequency as a function of year and traffic volume, both by track type and by accident cause. Third, collision severity (measured by the average number of railcars derailed per train collision) is estimated on the basis of historical accident data. Fourth, future train collision frequency and severity are projected on the basis of the safety trends of 2000 to 2014. Finally, the paper discusses the implications of the analyses with respect to research and practice.

DEFINITIONS

Collision

According to 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 main line and is struck by an approaching train. The FRA REA database includes 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. 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, that are expected to occur on the entity during a specified period (*18*). Liu provides a theoretical explanation for this definition from a stochastic-process perspective (*11*). 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 by using multivariate regression techniques, which are detailed later.

Risk

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

The definitions in this section form the basis of the following statistical analyses. Analysts may choose to use alternative definitions to fit their available data and questions of interest. In such cases, they can adapt this paper's methodology accordingly.

STATISTICAL METHODOLOGY FOR ANALYSIS OF TRAIN COLLISION RISK

Model Development

The FRA REA database records four types of track: main, siding, yard, and industrial. 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. presents an analysis of the frequency and severity of train derailments by the type of track and type of accident by using data from 2000 to 2010 (*5*). Table 1 shows an updated analysis that includes more recent data. In this paper, collisions include those between freight trains, as well as those between freight trains and nontrain consists (e.g., maintenance vehicles, cut of cars, locomotives).

The analysis shows that collisions accounted for approximately 6% of total accidents and 4% of the total number of railcars derailed on Class I freight railroads. Derailment was the most common type of

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

	Results by Type of Accident				
Track Type	Derailments	Collisions	Highway–Rail	Other	Total
Number of Freight Train Accidents					
Main line	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					
Main line	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					
Main line	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. When multiple trains are involved in a collision, the total number of railcars derailed from all the trains is counted as the collision severity.

accident on each track type, accounting for 72% of train accidents and 94% of the total number of railcars derailed across all types of tracks. On main tracks, the frequency of freight train collisions is about one-fifth that at grade crossings (429 versus 1,929). However, on main lines, the average number of railcars derailed per collision is nine times as great at grade crossing (4.3 versus 0.5). Of all highway–rail grade-crossing incidents, 99% 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. found that many grade-crossing incidents exceeded the FRA reporting threshold for monetary damages but did not result in a derailment (Table 1) (*14*).

A statistical model was developed to correlate frequency of train collisions with influencing factors. On the basis of data from the FRA REA database, this paper focuses on two potential affecting factors: year and annual traffic exposure. The year variable tests whether a temporal change occurred in the frequency of train collisions given traffic exposure. The traffic exposure variable describes whether and how the count of train collisions varies with traffic volume in a given year. A basic model structure is shown in Equation 1:

$$\mu_i = \exp(\alpha + \beta \times T_i + \theta \times M_i) M_i \tag{1}$$

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, and
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 rewritten as Equation 2.

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

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

The next step is to estimate the unknown parameters on the basis of 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 elsewhere (*23*). This paper starts with negative binomial regression. Where 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 that represents the statistical significance of a predictor variable by using the Wald test (*23*). A generally acceptable rule is that if a predictor variable has a *P*-value smaller than 5%, the variable is significant. The analysis found that the parameter coefficient for the variable year is significantly negative ($\beta = -.0414$, $P < .0001$); this result indicates a significant temporal decline in train collision rate given traffic exposure (Table 2).

Goodness of fit of a negative binomial model can be evaluated by using a statistical criterion called deviance. Statistical theory dictates that deviance asymptotically follows a chi-square (χ^2) distribution (*23*). On the basis of 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%, the model will appear to have an adequate fit to the empirical data. In the example, the deviance is 13.3, the degrees of freedom are 12, and the corresponding *P*-value = .35 (Table 2). On the basis of the fitted parameters, the expected frequency of U.S. freight train collisions is estimated by using Equation 3.

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

Equation 3 is written in a mathematically equivalent way as Equation 4.

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

Equation 4 estimates collision rate of Class I railroad freight trains on all types of tracks. The analysis shows that the expected annual collision rate declined between 2000 and 2014. In addition to the temporal change, collision rate increased with traffic volume, all else being equal. Table 3 compares the observed and the predicted number of freight train collisions by using the negative binomial regression model described earlier.

In addition to deviance, another common goodness-of-fit test is the χ^2 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, and
n = sample size (number of years).

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

Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > Chi-Square
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	.0438

NOTE: Deviance = 13.3; degrees of freedom = 12; P = .35.

TABLE 3 Empirical Versus Predicted Frequency of Freight Train Collisions, All Tracks, Class I Freight Railroads, 2000–2014

Year	Total Train Miles (millions)	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

On the basis of Table 3, $\chi^2 = 17.7$. The corresponding *P*-value is .22 (degrees of freedom = 14). Therefore, the analysis indicates that the estimated collision frequency reasonably matches the observed count. Both the deviance test and the χ^2 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 main line and yard tracks.

Major Collision Causes

The FRA REA database records more than 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 of the FRA subgroups was developed by Arthur D. Little, Inc. (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 but are more fine grained for certain causes and thereby allow greater resolution in some cases. For example, the FRA grouping combines broken rails, joint bars, and rail anchors in the same subgroup, whereas the ADL grouping distinguishes between broken rail and joint bar defects. These groups were used to analyze cause-specific collision frequency and severity. The ADL grouping of accident causes might not be the only grouping approach. Furthermore, the same cause may fall into multiple groups. Therefore, if analysts use a different scheme for grouping accident causes, the analyses should be adapted accordingly.

According to Figure 1, on main tracks (including siding tracks), 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),violation of switching rules is the top cause. A detailed breakdown of the cause codes within each ADL cause group can be found elsewhere (*24*).

By using the negative binomial regression described earlier, a cause-specific model for frequency of train collisions is developed (Table 4). All the models presented in Table 4 have been validated through the deviance and χ^2 tests. Each model represents the best fit to the empirical data.

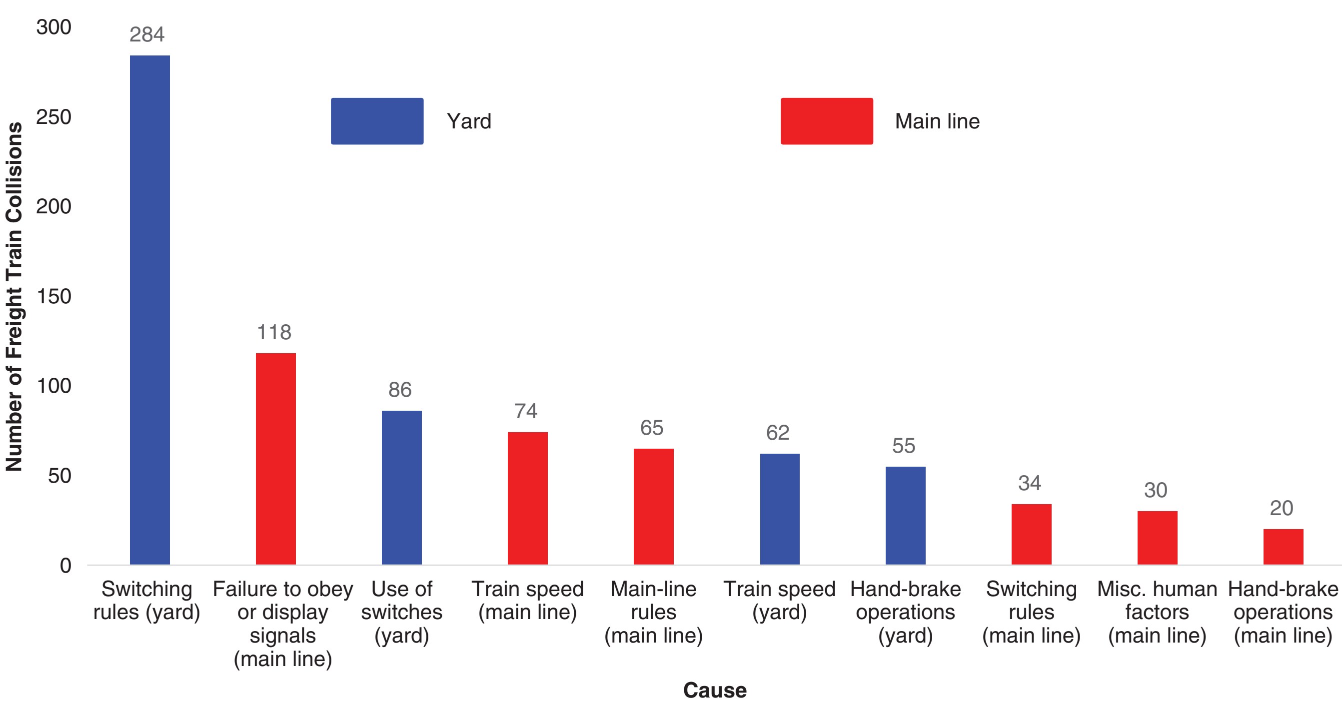

FIGURE 1 Top ADL collision cause group on main line and yard tracks, Class I freight railroads, 2000–2014 (misc. = miscellaneous).

TABLE 4 Selected Regression Models for Frequency of Freight Train Collisions

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

COLLISION SEVERITY

In addition to collision frequency, severity is another important element in the risk analysis of train collisions. 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 prior work (*1, 4–8*). On average, main line collisions tend to derail more railcars than collisions on yard and industrial tracks. A Wald–Wolfowitz runs test was used to determine whether collision severity follows any significant temporal trend (*25*). This statistical test checks whether a data set results from a random process. When the *P*-value in the test is larger than .05, one may conclude that no statistically significant trend for average annual collision severity. Table 5 shows that the severities for all the selected cause groups have no significant temporal trend. The yearly severity fluctuation is largely from random variations. Therefore, the following risk analysis uses the average of collision severity by type of track and accident cause (Table 5).

ANALYSIS OF COLLISION RISK

"Train collision risk" can be defined as the product of collision frequency and severity, as shown in Equation 6:

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

where

R = estimated annual collision risk,
F = estimated annual collision frequency, and
S = estimated collision severity (number of railcars derailed per train collision).

Both the estimated frequency and severity are subject to statistical uncertainty. Correspondingly, uncertainty is associated with the risk estimator. The variance of the risk estimator can be calculated by using the following equation and 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*):

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

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

		Railcars Derailed per Collision, by Cause				
		Main Line			Yard	
Year	All Tracks, All Causes	Failure to Obey or Display Signals	Violation of Train Speed Rules	All Causes on Main Lines	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	23.01	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	.11	.09	.16	.25	.20	.20

where

$\mathrm{var}(R)$ = variance of collision risk estimator,
$\mathrm{var}(F)$ = variance of collision frequency estimator,
$\mathrm{var}(S)$ = variance of collision severity estimator,
$E(F)$ = the estimator of collision frequency, and
$E(S)$ = the estimator of collision severity.

Furthermore, the 95% confidence interval of the collision risk estimator ($\mathrm{CI}_{95\%}(R)$) is

$$\mathrm{CI}_{95\%}(R)=\left[R-1.96\sqrt{\mathrm{var}(R)},\ R+1.96\sqrt{\mathrm{var}(R)}\right] \quad (8)$$

On the basis of Equations 6 to 8, the estimated annual collision risk and its 95% confidence interval are calculated by using the average traffic volume between 2000 and 2014 (Table 6). For example, if the traffic volume in 2017 is 533 million train miles on Class I main lines, the projected collision frequency is 19, with a 95% confidence interval of 14 to 25. This projection entails a 95% 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 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 main line risk is more sensitive to traffic than the risk on yard track. For example, if the traffic volume increases from 533 million to 549 million train miles on main lines in 2015 (3% increase), the estimated collision risk increases from 95 to 104, a 9% increase (Scenarios 2 and 3 in Table 6). By contrast, the collision risk on yard track increases 3%. The reason for the difference is that main line 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.

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

DISCUSSION OF CONTRIBUTIONS

This section discusses the contributions of this study with respect to the literature and practice.

Contributions to Literature

Because FRA has been collecting train accident data since the 1970s, researchers are able to examine historical accident data, to discover useful information, and to propose risk-informed decisions. Compared with the highway safety community, in which statistical modeling of accident data is normative for research and policy making, the U.S. railroad sector has used much less statistical modeling work. While

TABLE 6 Projected Collision Frequency, Severity, and Risk, 2015–2017, Class I Freight Railroads, by Traffic Volume

	Main Line Collision Estimates									
		Collision Frequency[a,b]			Collision Severity (number of cars derailed per collision)[c,d]			Collision Risk (total number of cars derailed)[b,e]		
Year	Main Line 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
Scenario 1. Baseline Traffic Decreases by 3%										
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
Scenario 2. Baseline Traffic Volume (average traffic volume, 2000–2014)										
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
Scenario 3. Baseline Traffic Volume Increases by 3%										
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

[a] The 95% upper bounds and lower bounds of collision frequency were generated in a negative binomial regression model by using GENMOD procedure in SAS statistical software. It accounts for the variance-covariance matrix of parameter coefficient estimators. The algorithm for generating the confidence interval of the predicted value is presented in SAS Manual, Chapter 37, GENMOD procedure.
[b] Estimated collision frequency and risk values were rounded to the nearest integers.
[c] The 95% upper bounds and lower bounds of the estimated collision severity (number of cars derailed) were developed on the basis of sample means and sample standard errors of the annual collision severity between 2000 and 2014.
[d] It is assumed that collision severity (number of railcars derailed) does not vary with traffic volume.
[e] The risk interval was developed by using Equation 7.

most existing railroad safety studies have concentrated on derailments or grade-crossing incidents, very limited statistical research has been spent on analysis of train collision risks. 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 few decades has been a statistical phenomenon called "regression to the mean," which refers to the tendency of a random variable that deviates from the mean to return to normal given that nothing else has changed (*18*). In the context of rail safety, regression to the mean implies that a high accident rate in 1 year may be followed by a low rate in the next year because of random fluctuation, even if no actual safety change occurred (*11*). As Liu mathematically proves, regression to the mean is inherent in any empirical train accident data and must be addressed through statistical approaches to understand the true safety trend (*11*). The collision analysis presented in this paper provides a step-by-step procedure for identifying the data-driven safety performance function in the railroad industry and accounts for random fluctuations in accident occurrence and severity. The general approaches and procedures here can be adapted to other types of train accidents and severities.

Contributions to Practice

The Rail Safety Improvement Act requires railroads to adopt risk-based approaches to ensure operational safety. However, no definitive, normative, practical methodological framework is available to guide the process of risk assessment. By contrast, FHWA publishes an extensive manual to guide the statistical modeling of vehicle crash data that is based on decades of research. The *Highway Safety Manual* provides practitioners with both easy-to-understand tutorials on the basic concepts of transportation statistics and toolboxes for statistically analyzing and interpreting accident data (*27*). The author hopes to use this paper, in conjunction with its sequent studies, to promote industry- and academiawide dialogues to develop a railroad safety analysis manual, just as the highway sector has done over the last 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 that technology's safety benefit. Ultimately, the railroad research community could develop a data-driven guideline for optimal safety investment.

CONCLUSIONS

This paper develops a statistical risk analysis of freight train collisions in the United States on the basis of data from 2000 to 2014. The analysis shows a temporal decline in collision rate on both main line 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 the risk of freight train collisions and enable a data-driven assessment of the safety effectiveness of certain accident prevention strategies.

Yard Collision Estimates									
	Collision Frequency[a,b]			Collision Severity (number of cars derailed per collision)[c,d]			Collision Risk (total number of cars derailed)[b,e]		
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
65.5	27	22	32	1.5	1.3	1.7	41	31	50
65.5	26	21	31	1.5	1.3	1.7	39	30	48
65.5	24	20	30	1.5	1.3	1.7	36	27	45
67.53	27	23	33	1.5	1.3	1.7	41	31	50
67.53	26	21	32	1.5	1.3	1.7	39	29	49
67.53	25	20	31	1.5	1.3	1.7	38	28	47
69.56	28	23	34	1.5	1.3	1.7	42	32	52
69.56	27	22	33	1.5	1.3	1.7	41	31	50
69.56	26	21	32	1.5	1.3	1.7	39	29	49

NEXT STEPS

First, the methodology will be adapted to account for other collision consequences, such as casualties, property damage, or environmental impact. These modifications will enable a comprehensive evaluation of train collision risk. Second, the methodology can be applied to passenger or rail transit accident analyses. Third, recent concern has been raised about the risks of transporting crude oil. The collision frequency and severity prediction models will be incorporated into a model to analyze hazardous materials risk for estimating the size of the population affected by a potential crude oil release. Fourth, a framework for before–after safety evaluation can be developed to assess how certain technologies affect collision risk. Finally, this paper focuses on statistics on freight train accidents in the United States. Other researchers studied train accidents in Canada (*3, 28, 29*) and Europe (*22*). Although a full comparison of rail safety statistics from the United States and those from other countries is beyond the scope of this paper, adapting the statistical methodology to other rail systems in future research may provide interesting and useful information.

ACKNOWLEDGMENTS

The author received funding in part from the Department of Civil and Environmental Engineering, the Center for Advanced Infrastructure and Transportation, and the School of Engineering, all at Rutgers University. The author thanks Shuang Xu and Yanlei Wang for their assistance.

REFERENCES

1. Barkan, C. P. L., C. T. Dick, and R. Anderson. Railroad Derailment Factors Affecting Hazardous Materials Transportation Risk. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1825,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 64–74.
2. Anderson, R. T., and C. P. L. Barkan. Railroad Accident Rates for Use in Transportation Risk Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1863,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 88–98.
3. Bagheri, M., F. F. Saccomanno, S. Chenouri, and L. P. Fu. 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., C. P. L. Barkan, and M. R. Saat. Analysis of Derailments by Accident Cause: Evaluating Railroad Track Upgrades to Reduce Transportation Risk. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2261,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 178–185.
5. Liu, X., M. R. Saat, and C. P. L. Barkan. Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2289,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 154–163.
6. Liu, X., M. R. Saat, and C. P. L. Barkan. Integrated Risk Reduction Framework to Improve Railway Hazardous Materials Transportation Safety. *Journal of Hazardous Materials,* Vol. 260, 2013, pp. 131–140.
7. Liu, X., M. R. Saat, and C. P. L. Barkan. Safety Effectiveness of Integrated Risk Reduction Strategies for Rail Transport of Hazardous Materials. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2274,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 102–110.
8. Liu, X., M. R. Saat, X. Qin, and C. P. L. Barkan. Analysis of U.S. Freight-Train Derailment Severity Using Zero-Truncated Negative Binomial Regression and Quantile Regression. *Accident Analysis and Prevention,* Vol. 59, 2013, pp. 87–93.
9. Liu, X., M. R. Saat, and C. P. L. Barkan. Probability Analysis of Multiple-Tank-Car Release Incidents in Railway Hazardous Materials Transportation. *Journal of Hazardous Materials,* Vol. 276, 2014, pp. 442–451.
10. Liu, X., A. Lovett, C. T. Dick, M. R. Saat, and C. P. L. Barkan. Optimization of Ultrasonic Rail-Defect Inspection for Improving Railway Transportation Safety and Efficiency. *Journal of Transportation Engineering,* Vol. 140, No. 10, 2014.
11. Liu, X. Statistical Temporal Analysis of Freight Train Derailment Rates in the United States: 2000 to 2012. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2476,* Transportation Research Board, Washington, D.C., 2015, pp. 119–125.
12. Austin, R. D., and J. L. Carson. An Alternative Accident Prediction Model for Highway–Rail Interfaces. *Accident Analysis and Prevention,* Vol. 34, 2002, pp. 31–42.
13. Saccomanno, F. F., P. Y.-J. Park, and L. Fu. 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., M. R. Saat, and C. P. Barkan. Analysis of Factors Affecting Train Derailments at Highway–Rail Grade Crossings. Presented at 91st Annual Meeting of the Transportation Research Board, 2012.
15. *Guideline for Accident Report Preparation.* FRA, U.S. Department of Transportation, 2011.
16. Nayak, P. R., D. B. Rosenfield, and J. H. Hagopian. *Event Probabilities and Impact Zones for Hazardous Materials Accidents on Railroads.* DOT/FRA/ORD-83/20. FRA, U.S. Department of Transportation, 1983.
17. *Class I Railroad Statistics.* Association of American Railroads, Washington, D.C., 2013.
18. Hauer, E. *Observational Before–After Studies in Road Safety.* Pergamon Press, Oxford, United Kingdom, 1997.
19. Kaplan, S., and B. J. Garrick. 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, Cambridge, United Kingdom, 2007.
24. *Risk Assessment for the Transportation of Hazardous Materials by Rail. Supplementary Report: Railroad Accident Rate and Risk Reduction Option Effectiveness Analysis and Data,* 2nd rev. Arthur D. Little, Inc., Cambridge, Mass, 1996.
25. Bradley, J. V. *Distribution-Free Statistical Tests.* Prentice Hall, New York, 1968.
26. Goodman, L. A. The Variance of the Product of *K* Random Variables. *Journal of the American Statistical Association,* Vol. 57, No. 297, 1962, pp. 54–60.
27. *Highway Safety Manual.* AASHTO, 2015. http://www.highwaysafetymanual.org/Documents/HSMP-1.pdf.
28. Saccomanno, F. F., and S. El-Hage. Minimizing Derailments of Railcars Carrying Dangerous Commodities Through Effective Marshaling Strategies. In *Transportation Research Record 1245,* TRB, National Research Council, Washington, D.C., 1989, pp. 34–51.
29. Saccomanno, F. F., and S. M. El-Hage. Establishing Derailment Profile by Position for Corridor Shipments of Dangerous Goods. *Canadian Journal of Civil Engineering,* Vol. 18, No. 1, 1991, pp. 67–75.

The author is responsible for the views and analyses in this paper.

The Standing Committee on Railroad Operational Safety peer-reviewed this paper.