Institution of MECHANICAL ENGINEERS

Journal of RAIL AND RAPID TRANSIT

Original Article

Proc IMechE Part F:
J Rail and Rapid Transit
0(0) 1–14

DOI: 10.1177/0954409718811742
journals.sagepub.com/home/pif
SAGE

Analysis of freight train collision risk in the United States

Tejashree Turla, Xiang Liu and Zhipeng Zhang

Abstract

Rail transportation is pivotal for the national economy. Despite being rare, a train accident can potentially result in severe consequences, such as infrastructure damage costs, casualties, and environmental impacts. An understanding of accident frequency, severity, and risk is important for rail safety management. In the United States, extensive prior research has focused on risk analyses of train derailments and highway–rail grade crossing accidents. Relatively less work has been conducted regarding train collision risk. The US Federal Railroad Administration identifies various accident causes, among which the authors of this study have analyzed the major collision causes. For each major accident cause, the authors have analyzed its resultant collision frequency, severity (in terms of damage cost or casualties), and correspondingly the risk, which is the combination of the frequency and severity. The analysis was based on train collision data in the United States from 2001 to 2015. This analysis focuses on freight trains in the United States, due to their immense traffic exposure. On the temporal scale, collision rate (the number of collisions normalized by traffic exposure) has an approximately 5% annual reduction. In terms of collision cause, failures to obey signals, overspeeds, and violations of mainline operating rules accounted for more collisions than other causes. Two alternative risk measures, namely the expected consequence and conditional value at risk, were used to evaluate the freight train collision risk on main tracks, accounting for both the average and worst-case scenarios. This collision risk analysis methodology may provide the US Department of Transportation and railroad industry with information and decision support for identifying, evaluating, and implementing cost-effective risk mitigation strategies.

Keywords

Railroad, train collisions, freight, risk, accident cause

Date received: 28 March 2018; accepted: 13 October 2018

Introduction

Rail is a safe mode of surface transportation. However, an accident may cause infrastructure and rolling stock damages, possible casualties and environmental impacts. Risk analysis is useful to understand the characteristics of historical accidents,[1] and thus to develop appropriate accident prevention strategies. Train accident risk can be defined as the probability distribution of its possible consequences (e.g. in terms of casualties or damage costs).[2] The United States has the largest freight rail network in the world, including over 140,000 mi of track. Most of the related prior literature has focused on either train derailments[3–6] or highway–rail grade crossing accidents,[7,8] with much less research on train collision risk in the U.S. This knowledge gap motivates the development of this paper.

The goal of this research is to analyze U.S. train collision data, and thus to develop a statistical risk model accounting for major train collision causes, accident frequency, and severity. The data used in this study came from the Federal Railroad Administration (FRA) of the US Department of Transportation (USDOT). The analysis focuses on freight train collisions on main line tracks during the period of 2001–2015. This paper contributes to the academia as follows:

- First, in addition to empirical analysis, the statistical risk analysis specific to the major causes causing train collisions is an original contribution of this research. While the methodology is particularly applied to train collision risk analysis, it can be generalized and adapted to other types of train

Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, Piscataway, USA

Corresponding author:
Xiang Liu, Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, CoRE 606, 96 Frelinghuysen Road, Piscataway, USA.
Email: xiang.liu@rutgers.edu

accidents as well. Compared to the empirical approach, statistical risk analysis accounts for the random fluctuation of accident frequency and severity and focuses on the trend.
- Second, it introduces the concept of spectral risk measure (SRM) in the field of railway safety analysis. SRMs are the risk measures that account for risk aversion, which in the railway safety is, considering the low-probability–high-consequence characteristics of train accidents. To our knowledge, this is the first study to analyze collision risk using conditional value at risk (CVaR) as the alternative risk measure. It is the weighted average of quantiles of a severity distribution within a desired confidence interval and is derived from the integration of simulation with statistical risk analysis.

Data

All the railroads operating in the United States (including Canadian railways that have subsidiaries in the U.S.) are required to report all accidents that exceed a monetary threshold of damages to infrastructure, rolling stock, and signals. FRA compiles these accident reports into the Rail Equipment Accident (REA) database, which records all REA data back to 1975. The FRA REA database[9] records railroad, accident type, location, accident cause, severity, and other information important for accident analysis and prevention. There are four types of tracks included in the FRA REA database, which are main track, siding track, yard track, and industry track, respectively. These track types are used for different operational functions and consequently have different associated accident types, causes, and consequences. Train accidents are categorized into derailment, collision, highway–rail grade crossing accident, and other less frequent types. FRA train accident cause codes are hierarchically organized and categorized into major cause groups, which are track, equipment, human factors, signal, and miscellaneous causes. Within each of these major cause groups, FRA organized individual cause codes into subgroups of related causes, which were refined by Arthur D. Little, Inc.[10] In addition, the REA database also contains accident severity information in terms of damage cost to infrastructure and rolling stock, casualties, and hazardous material cars releasing contents (if any). The FRA REA database has been used in numerous previous safety analyses with respect to train derailments (e.g. Barkan et al.,[3] Liu et al.[4–6]), hazardous materials release incidents,[3,11–13] and grade crossing collision incidents.[7,8] In addition to the safety data, each railroad also reports their monthly train-mile data to FRA through the Operational Safety Database. In this paper, we use both accident data and traffic exposure data to quantitatively evaluate freight train collision risk, in terms of collision rate per unit of traffic exposure, collision severity (either casualties or damage cost), and their combination via alternative risk measures such as the expected consequence or CVaR.

Freight train collision rate

This analysis includes all common types of train collisions (head-on collision, rear-end collision, side collision, raking collision, and broken train collision). This analysis focuses on the collisions between trains, excluding the consideration of the collisions between highway vehicles and trains at the interface of highway–rail grade crossings. Grade crossing incident has different accident characteristics and should be treated in separate analyses.[8] In the United States, passenger trains share most of the trackage with freight railroads, causing potential collisions between freight trains and passenger trains. However, as the historical data suggest, over 89% of train collisions occurred between freight trains (394 out of 444), with the passenger-train-to-freight train collisions (19) and passenger-train-to-passenger-train collisions (32) only accounting for 7% and 4% of overall collision frequency, respectively, from 2001 to 2015. Recognizing that different types of train collisions may have different accident characteristics and risks, we develop separate analyses for each of them. This paper will entirely focus on the 394 collisions between freight trains on main tracks due to the prevalence of this type of collisions in the United States.

This paper develops a negative binomial (NB) regression model specifically to analyze freight train collision rates on main tracks in the national scale. In the United States, location-centric traffic volume information is generally proprietary to freight railroads. Therefore, in this paper, we focus on nationwide aggregated accident count. The variables considered in our study include year and traffic exposure. Due to this scope, the NB model could be adequate. A similar use of the NB model for nationwide railway safety analysis was also seen in the literature (e.g. Evans[14] and Liu[2,15]). The NB model has also been widely used in accident rate analysis in highway transportation[16–22] and its basic framework is as follows

$$Y \sim Poisson(\lambda) \tag{1}$$

$$\lambda \sim Gamma\left(f, \frac{f}{m}\right) \tag{2}$$

$$m = \exp\left(\sum_{p=0}^{k} b_p X_p\right) M \tag{3}$$

where

$Y =$ observed number of collisions
$m =$ estimated number of collisions
$b_p = p$th parameter coefficient

$X_p = p$th explanatory variable
$M =$ traffic exposure (e.g. train-miles).
$f =$ inverse dispersion parameter

In this research, two predictor variables are considered, which are the year index and traffic volume for collision statistical analysis (equation (4)). The selection of these two variables is consistent with a prior study.[2] The year variable tests if there is a temporal change in the frequency of train collisions with a given traffic exposure. Similarly, the traffic exposure variable tests whether and how the number of train collisions may vary with the traffic volume in a given year. The results of the parameter coefficients in equation (4) are presented in Table 1 with a 95% confidence interval

$$\mu_i = \exp(\alpha + \beta \times T_i + \theta \times M_i)M_i \qquad (4)$$

where

$\mu_i =$ expected number of train collisions in ith year
$M_i =$ traffic exposure in ith year (e.g. billion train-miles)
$T_i =$ year index
$\alpha, \beta, \theta =$ parameter coefficients

The train collision rate is defined as the number of train collisions normalized by traffic exposure. With respect to this definition, equation (4) can be modified as follows

$$Z_i = \frac{\mu_i}{M_i} = \exp(\alpha + \beta \times T_i + \theta \times M_i) \qquad (5)$$

where
$Z_i =$ expected train collision rate per billion train miles in ith year

A similar model was used in previous studies[23,24] where train accident rate is assumed to be independent of traffic exposure ($\theta = 0$). To understand whether and how collision frequency may vary 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 parameter coefficients, α, β, and θ were estimated using the maximum likelihood method. The results of estimated frequency in each year are presented in Table 2, with the observed frequency, including the lower and upper bound values with a confidence interval of 95%. In the model (Table 1), the p-value of a parameter estimator which is obtained using the Wald

Table 1. Negative binomial regression of freight train collision frequency, 2001–2015.

Parameter	Estimate	Standard error	95% Confidence limits	p-value	
α	97.53	30.20	31.73	163.32	0.0072
β	−0.05	0.01	−0.08	−0.02	0.0075
θ	3.45	1.81	−0.49	7.39	0.0808

Table 2. Comparison of empirical and estimated frequency values by year.

		Collision frequency			
Year	Train-miles (billions)	Observed	Estimated	95% Upper bound of the estimator	95% Lower bound of the estimator
2001	0.54	38	32	39	29
2002	0.55	26	32	36	28
2003	0.56	31	32	37	28
2004	0.58	40	34	38	30
2005	0.60	54	36	40	32
2006	0.62	25	38	42	34
2007	0.59	30	31	35	27
2008	0.57	24	27	31	23
2009	0.48	14	16	20	12
2010	0.51	17	18	22	13
2011	0.52	16	18	22	14
2012	0.53	24	18	22	14
2013	0.54	21	18	22	14
2014	0.56	21	19	23	15
2015	0.53	13	16	20	11

test[25] denotes the statistical significance of the respective predictor variable. If a predictor variable has a p-value smaller than 5% then according to a commonly acceptable rule, the variable is considered significant. The analysis found that the parameter coefficient for the variable year (which represents the annual change of collision rate) is negative ($\beta = -0.05$, p-value $= 0.0075$); this result indicates a significant temporal decline in train collision rate (approximately an average of 5% annually) given traffic exposure (Figure 1). The p-value for the parameter coefficient of the variable traffic is 0.0808, indicating that freight train collision rate does not vary much with traffic exposure. In other words, for a given year, collision frequency increases linearly with traffic exposure.

The goodness of fit of this model is evaluated using a Chi-squared test (equation (6)), which assesses the relative difference between the estimated and observed values. If the p-value in the test is larger than 5%, the model will appear to have an adequate fit to the empirical data. On the basis of Table 2, the $\chi^2 = 19.93$ and the corresponding p-value is 0.13 (degrees of freedom $= 14$). This shows that the estimated collision frequency is close enough to the observed count. The empirical collision frequency lies within the 95% confidence interval of the estimated frequency

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

where

O_i = observed number of collisions in ith year

E_i = estimated number of collisions in ith year

n = sample size (number of years in the study period

The model can be used to predict the frequency for any year in the future by substituting the desired year variable and a known volume of annual traffic. Due to the availability of traffic and accident data in 2016 and 2017, the model-estimated values are compared with the respective empirical frequencies in Table 3. The empirical frequencies lie in the estimated 95% confidence interval.

Severity analysis

Severity is the measure of intensity of an impact and is a major factor in determining the collision risk. It can be measured using different proxy variables. Some previous studies used the number of railcars derailed as the severity proxy[3–5,26] and some used the number of casualties as the severity proxy.[14,27] For hazardous material transportation, the number of tank cars releasing was used to measure the severity.[2,12,13,28,29]

In some train collisions, even though there are no casualties, there exist damage costs due to infrastructure damages, etc. In this paper, we use the number of injuries, fatalities, and the damage costs as the proxy to measure collision severity. In order to check the randomness of a data set, a method called Wald–Wolfowitz runs test is used. The p-values for severity in injuries, fatalities, and damage costs obtained by runs test are 0.266, 0.266, and 0.095, respectively. Since these values are greater than 0.05, we conclude that there is no significant temporal trend of collision severity in the study period. Therefore,

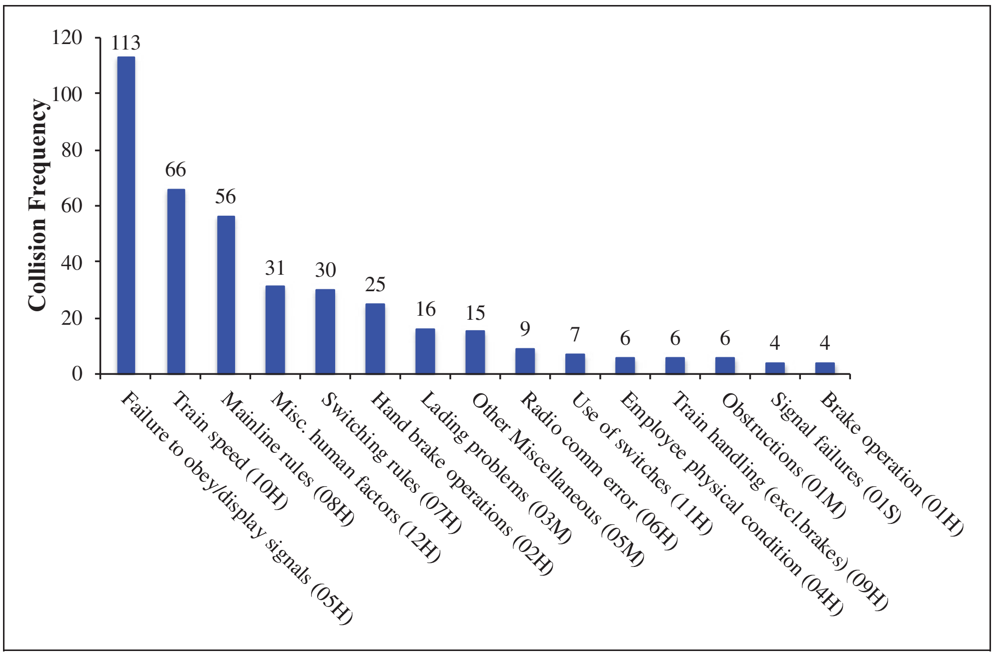

Figure 1. Number of collisions by cause, U.S. mainlines, 2001–2015.

Table 3. Predicted collision frequency on US freight mainlines, by traffic volume.

Year	Train-miles (in billions)	Collision frequency			
		Empirical	Mean estimate	95% Upper bound	95% Lower bound
2016	0.489	9	12	16	8
2017	0.498	9	12	16	8

Table 4. Different measures of severity per collision, by year (2001–2015).

Year	Number of collisions	Injuries per collision	Fatalities per collision	Damage costs per collision (million $)
2001	38	0.87	0.11	0.90
2002	26	1.00	0.04	0.90
2003	31	0.74	0.00	0.54
2004	40	2.58	0.20	0.68
2005	54	1.63	0.09	0.67
2006	25	1.00	0.00	0.56
2007	30	0.30	0.07	0.32
2008	24	1.04	0.00	0.47
2009	14	0.57	0.00	0.42
2010	17	0.94	0.06	0.40
2011	16	0.69	0.25	0.41
2012	24	0.42	0.17	0.85
2013	21	1.19	0.05	1.37
2014	21	0.76	0.10	1.01
2015	13	1.15	0.00	0.57
Average	26	0.99	0.07	0.67
Std. error	2.86	0.14	0.02	0.07
p-value in runs test		0.266	0.266	0.095

we can use the average (mean) number of injuries, fatalities, and damage costs to represent the overall collision severity.

Table 4 shows the severity measures in terms of injuries, fatalities, and damage costs for each year. The reported damage costs in each year have been adjusted to 2016 monetary values considering inflation. On average, a mainline freight train collision results in one injury or $0.67 million of monetary damage cost to infrastructure and rolling stock.

Causal analysis

More than 300 accident causes have been recognized by the FRA and recorded in the REA database. There are 16 causes (related to human factor, signals, other or miscellaneous cause groups) associated with the collisions on U.S. mainlines. Figure 1 (includes the collisions with zero injuries or fatalities) shows the number of collisions caused due to each cause group except one collision in brake operations (other) cause group (03H). Similarly, Figure 2 shows the average severity in terms of injuries, fatalities, and damage costs per collision due to each cause group. The frequency analysis shows that the failure to obey or display signals (05H), violation of train speed rules (10H), and violation of mainline operating rules (08H) are the top three collision causes. The individual cause codes and explanations are shown in Appendix 1 (ADL, 1996).[10]

Collision frequency by major causes

Using the NB regression model discussed earlier in the "Severity analysis" section, accident-cause-specific collision rate models are developed. The collision rates due to failure to obey or display signals (05H) and violation of mainline rules (8H) show statistically significant reduction in the study period. By contrast, there is no temporal trend of collision rate due to violation of train speed rules (10H) given the traffic exposure (Table 5). It can also be seen that the parameter for traffic exposure (θ) is insignificant for the collisions due to failure to obey or display signals (05H) or violation of mainline rules (08H) (θ>0.05). However, the collision rate due to violation of train

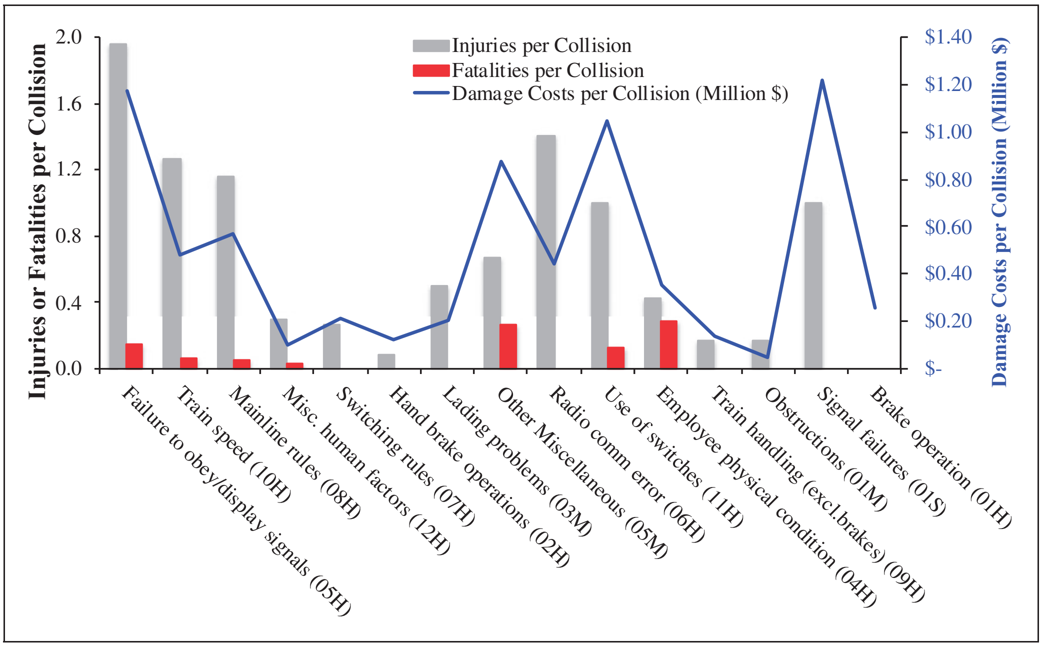

Figure 2. Collision severity by cause, U.S. mainlines, 2001–2015.

Table 5. Regression model for cause-specific collision frequency on mainlines.

Accident cause	Collision frequency (μ_i) by year (T_i) and traffic exposure measured by billion train-miles (M_i)	p-Value (from Chi-squared goodness-of-fit test)
Failure to obey or display signals	$\mu_i = \exp(140.33 - 0.07 \times T_i)M_i$	0.803
Train speed	$\mu_i = \exp(-3.08 + 8.99 \times M_i)M_i$	0.292
Mainline rules	$\mu_i = \exp(94.36 - 0.05 \times T_i)M_i$	0.932

speed rules increases when traffic volume increases. Table 4 shows the "final" regression model for each cause.

The goodness of fit of an NB model can be evaluated using a Chi-squared test. Since the p-value is greater than 0.05 for all the three major causes, the prediction reasonably reflects the empirical data according to the goodness-of-fit test. The empirical and model-based estimated values are presented in Table 6.

Collision severity of major causes

We further analyze the severity of the top three causes. We consider the mean values of injuries, fatalities, and damage costs per collision as the respective severity measure. Table 7 shows that failure to obey or display signals (05H) not only has the highest frequency (Table 6) but also leads to the highest number of injuries (1.6 per collision), fatalities (0.1 per collision), and damage cost ($1.23 million per collision) than other major collision causes. Compared to the violation of mainline rules (08H), the frequency of collisions caused by violation of train speed rules (10H) is higher but has a slightly lower severity (in terms of both injuries and damage cost).

Analysis of freight train collision risk

FN curve

In order to have a better understanding of the collision risk, we illustrate the *FN* curves, which can visually compare alternative risks.[30]

F (the vertical axis) represents the cumulative frequency of events that caused *N* or more casualties per collision (the horizontal axis). For convenience, we combined injuries and fatalities and grouped into casualties in the curves as zero fatality values cannot show up on logarithmic scale. Also, we focus on analyzing the collisions with at least one casualty for the same reason. Figure 3 presents the annual number of collisions, with *N* or more casualties per train collision, by major causes, in 2001–2015. It shows that failure to obey or display signals (05H) has a higher relative risk, due to its resultant higher collision frequency and severity. This accident cause involves situations such as automatic block signal, signal

Table 6. Observed and estimated freight train collision frequencies for the top three causes.

	Failure to obey or display signals (05H)		Violation of train speed rules (10H)		Violation of mainline rules (08H)	
Year	Observed	Estimated	Observed	Estimated	Observed	Estimated
2001	10	11	6	3	6	5
2002	9	10	1	4	4	4
2003	8	10	3	4	2	4
2004	14	10	6	5	6	4
2005	14	9	13	6	7	4
2006	7	9	7	7	3	4
2007	8	8	3	5	5	4
2008	7	7	7	4	3	3
2009	5	6	2	2	3	3
2010	4	6	3	2	4	3
2011	9	5	3	3	1	3
2012	4	5	3	3	3	3
2013	6	5	2	3	4	3
2014	3	5	5	4	3	3
2015	5	4	2	3	2	2
Average	8	7	4	4	4	3

equipment damages, inability to give or receive hand signals and flagging signals, etc. A detailed breakdown of the cause codes and explanations within each ADL cause group can be found in ADL.[10] Violation of mainline rules (08H) typically resulted in a relatively lower severity in terms of casualties (fewer than 10 casualties per collision). Similar to the failure to obey signals, the violation of train speed rules (10H) also has the potential to result in a large number of casualties.

Annual risk

According to the law of total expectation, if N is the total number of collisions with X_{ij} injuries in the jth accident, then the risk can be estimated using the following equation, given that both N (collision frequency) and X_{ij} (collision severity) are random variables

$$E\left[\sum_{j=1}^{N} X_{ij}\right] = E\left[E\left[\sum_{j=1}^{N} X_{ij}|N\right]\right] = E\left[NE\left[X_{ij}\right]\right] = E[N]E[X] \tag{7}$$

Equation (7) shows that the annual risk can be expressed as a product of expected accident frequency and severity. This is also called as the ''expected consequence'' risk measure.[2] The expected accident frequency can be estimated using the NB regression model described above, and the expected accident severity can be approximated by the sample mean. Because there is no significant temporal trend of collision severity in the study period, we use the 15-year average severity.

The estimated annual risk for each major cause in Table 8 is the product of their respective estimated frequency (Table 6) and overall average severity (Table 7). For example, the estimated frequency of accidents due to violation of train speed rules (10H) in 2001 is 3 and average severity is \$0.56 million damage cost; therefore, the estimated risk in damage cost for that year is 1.69 (3×0.56). The results indicate that the average annual risk due to failure to obey or display signals accounts for almost half the risk due to all causes and is three times the risk due to violations of train speeds. The collision risk due to violation of mainline operating rules is similar to that due to violation of train speed rules.

Alternative risk measures

The major limitation of using mean as the risk measure is that it fails to account for the extreme characteristics of the accident with low probabilities but high consequences. For example, almost 60% of the freight train collisions in the 15-year study period had zero casualties. While most collisions resulted in less than seven casualties, two collisions led to 47 and 82 casualties forming a long tail distribution (Figure 4). To address this ''heavy-tail'' effect, the prior literature used risk measures such as value at risk (VaR) or CVaR as alternative risk measures.[31,32]

VaR is defined as the loss level that will not be exceeded with a certain confidence level during a certain period of time. VaR at a confidence level of $\alpha \in (0, 1)$ of a variable X (casualties or damage costs) is the α-quantile of the collision distribution with respect to the frequency

$$VaR_\alpha(X) = q_\alpha(\mathrm{X}) = \inf\{x : P(X \leqslant x) \leqslant 1 - \alpha\} \tag{8}$$

Table 7. Severity of top three causes in terms of casualties and damage cost per collision.

Year	Failure to obey or display signals (05H) Injuries	Fatalities	Damage costs (million $)	Train speed (10H) Injuries	Fatalities	Damage costs (million $)	Mainline rules (08H) Injuries	Fatalities	Damage costs (million $)
2001	2.28	0.10	2.37	0.17	0.00	0.40	0.33	0.00	0.10
2002	1.32	0.00	1.08	1.00	0.00	0.04	1.75	0.25	2.55
2003	1.24	0.00	1.40	0.67	0.00	0.19	0.50	0.00	0.02
2004	6.51	0.35	1.29	0.17	0.00	0.14	1.50	0.17	0.92
2005	0.71	0.28	1.50	4.08	0.08	0.50	1.14	0.00	0.64
2006	1.70	0.00	1.32	0.71	0.00	0.37	1.67	0.00	0.19
2007	0.62	0.25	0.71	0.00	0.00	0.03	0.60	0.00	0.30
2008	1.13	0.00	1.22	0.86	0.00	0.20	3.33	0.00	0.20
2009	1.59	0.00	1.10	0.00	0.00	0.04	0.00	0.00	0.06
2010	0.74	0.25	0.66	0.00	0.00	0.23	2.25	0.00	0.83
2011	0.99	0.44	0.66	0.67	0.00	0.04	0.00	0.00	0.15
2012	0.25	0.00	0.52	1.00	1.00	5.07	1.33	0.00	0.23
2013	2.81	0.00	3.08	0.50	0.00	0.06	1.00	0.25	0.18
2014	1.65	0.00	1.06	1.00	0.00	0.91	0.33	0.00	0.08
2015	0.79	0.00	0.48	1.00	0.00	0.21	4.50	0.00	2.35
Average	1.62	0.11	1.23	0.79	0.07	0.56	1.35	0.04	0.59
Std. error	0.4	0.0	0.2	0.3	0.1	0.3	0.3	0.0	0.2

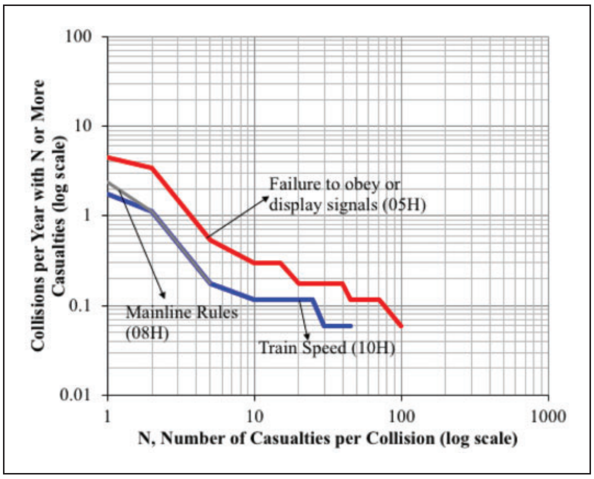

Figure 3. *FN* curve for cause-specific freight train collisions.

CVaR, or sometimes called as expected shortfall, in short, is the weighted average of all outcomes exceeding the confidence interval of a data set sorted from worst to best. Simply, CVaR of the collision risk is the average of all the number of casualties or damage costs that are more than $\alpha \in (0, 1)$

$$CVaR_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(X)\mathrm{d}u \tag{9}$$

Many previous studies prefer CVaR to VaR, due to its coherency.[33,34] This is because VaR does not reveal anything about the magnitude of losses exceeding the VaR limit. A risk measure is said to be coherent if it exhibits monotonicity, subadditivity, positive homogeneity, and translational invariance,[34] which are the properties followed by CVaR. Therefore, in the further analysis, CVaR is chosen as an alternative risk measure to mean. A confidence interval of 95% has been adapted to obtain the mean of 5% worst-case collisions with severity in terms of casualties or damage costs

$$R_i = CVaR_{95\%}\left(\sum_{j=1}^{N} X_{ij}\right) \tag{10}$$

where

$i = 1$, using number of casualties as collision severity metric
$= 2$, using damage cost as collision severity metric
$R_i =$ annual collision risk based on the specific severity metric used
$N =$ number of collisions in a specific year
$X_{ij} =$ collision severity (e.g. casualties or damage cost)

Annual risk in alternative measure

CVaR is typically not analytically tractable; therefore, Monte Carlo simulation[35] is used in this study. First, several probabilistic distributions are used to fit the empirical distribution of collision severity (in either damage cost or casualties), using a statistical tool called EasyFit (version 2017). Then, the "best"

Table 8. Estimated annual risk due to the three major causes and all causes combined, 2001–2015.

	Failure to obey or display signals (05H)			Violation of train speed rules (10H)			Violation of mainline rules (08H)			All causes		
Year	Risk in injuries	Risk in fatalities	Risk in damage cost	Risk in injuries	Risk in fatalities	Risk in damage cost	Risk in injuries	Risk in fatalities	Risk in damage cost	Risk in injuries	Risk in fatalities	Risk in damage cost
2001	17.8	1.2	13.53	2.4	2.0	1.69	6.7	0.2	2.93	31.9	2.4	21.85
2002	16.2	1.1	12.30	2.4	2.0	1.69	5.4	0.2	2.35	31.7	2.4	21.75
2003	16.2	1.1	12.30	3.2	2.0	2.25	5.4	0.2	2.35	32.2	2.4	22.11
2004	16.2	1.1	12.30	3.9	2.0	2.81	5.4	0.2	2.35	34.2	2.6	23.43
2005	14.6	1.0	11.07	3.9	2.0	2.81	5.4	0.2	2.35	34.6	2.6	23.70
2006	14.6	1.0	11.07	6.3	2.0	4.49	5.4	0.2	2.35	38.5	2.9	26.39
2007	13.0	0.9	9.84	3.9	2.0	2.81	5.4	0.2	2.35	30.4	2.3	20.85
2008	11.4	0.8	8.61	3.2	2.0	2.25	4.0	0.1	1.76	26.2	2.0	17.96
2009	9.7	0.7	7.38	1.6	2.0	1.12	4.0	0.1	1.76	15.9	1.2	10.87
2010	9.7	0.7	7.38	1.6	2.0	1.12	4.0	0.1	1.76	17.6	1.3	12.06
2011	8.1	0.6	6.15	2.4	2.0	1.69	4.0	0.1	1.76	18.0	1.4	12.33
2012	8.1	0.6	6.15	2.4	2.0	1.69	4.0	0.1	1.76	18.1	1.4	12.41
2013	8.1	0.6	6.15	3.2	2.0	2.25	4.0	0.1	1.76	18.2	1.4	12.45
2014	8.1	0.6	6.15	3.2	2.0	2.25	4.0	0.1	1.76	18.6	1.4	12.72
2015	6.5	0.4	4.92	2.4	2.0	1.69	2.7	0.1	1.17	15.3	1.1	10.46
Average	11.9	0.8	9.02	3.0	2.0	2.17	4.7	0.2	2.03	25.4	1.9	17.42

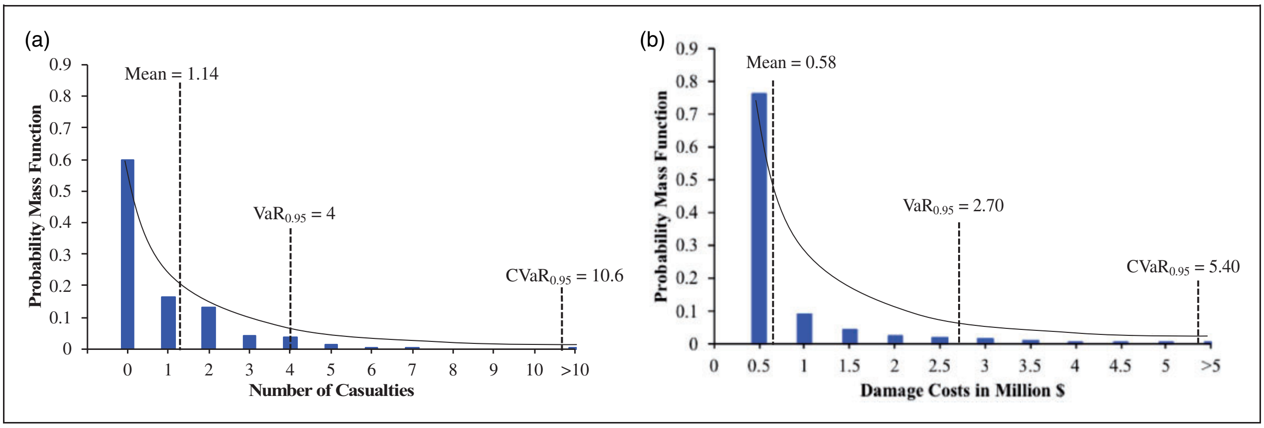

Figure 4. Distribution of severity per freight train collision, 2001–2015. (a) Casualties and (b) damage cost. CVaR: conditional value at risk; VaR: value at risk.

model is selected based on the Kolmogorov–Smirnov test. For the 15-year study period, the gamma distribution and log-normal distribution are selected as the ''best'' fitted distributions for casualties and damage cost, respectively. Then N different samples (train collisions) in a specific year are generated, where N follows a NB distribution as discussed in ''Freight train collision rate'' section. Since the severity does not follow a trend, the 15-year average of accident consequences (either casualties or damage cost) is calculated among N different samples that were obtained before and repeated for 100,000 iterations. Finally, the simulated CVaR is calculated for any given year.

Figures 5 and 6 present the annual train collision risk due to each major cause using different severity metrics. The $CVaR_{95\%}$ gives the average severity of the worst 5% of collisions, which is unsurprisingly higher than the mean value. The average annual freight train collision risk due to failure to obey or display signals is 14 casualties or \$8.75 million damage cost in the study period. In comparison, the average of the worst 5% of freight train collisions causes 81 casualties or \$24.95 million damage cost annually for this particular cause. From the p-value in the runs test (< 0.05) for collision risk due to failure to obey or display signals and violation of

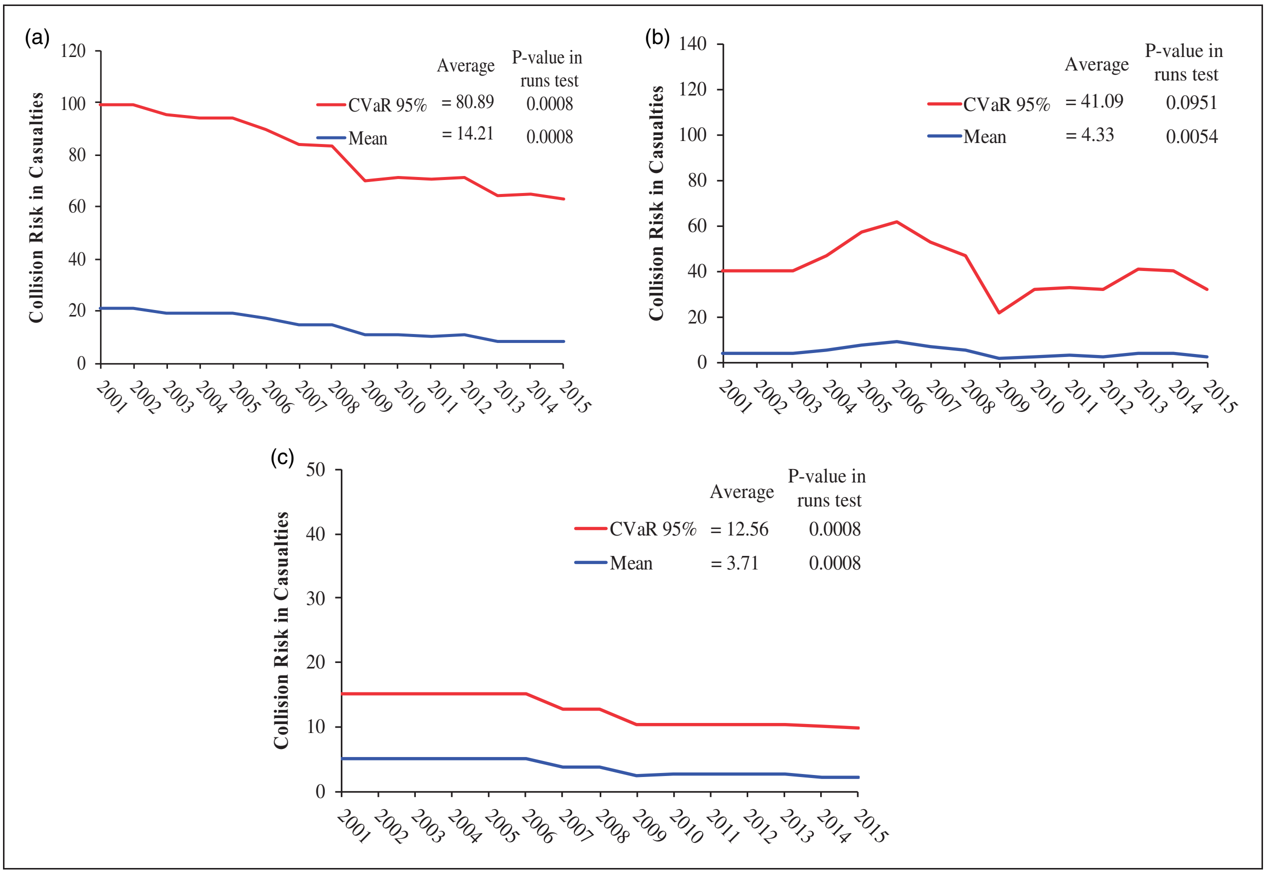

Figure 5. Freight train collision risk in terms of casualties for major causes, 2001–2015. (a) Failure to obey or display signals (05H), (b) violation of train speed rules (10H), and (c) violation of mainline rules (08H). CVaR: conditional value at risk.

mainline rules both in casualties and damage costs, it can be concluded that the risk is not random and is actually decreasing over the study period.

Discussion of collision mitigation with positive train control (PTC)

The analysis above shows that that failure to obey signals (05H) and violation of train speed rules (10H) are two major causes that may lead to high-frequency–high-severity train collisions. PTC can prevent certain accidents due to these causes.[36] In particular, PTC is designed to prevent train accidents attributable to excessive speeds, improper train movements, and other human error, by slowing or stopping trains automatically.[37] To achieve these functions, PTC systems use a combination of communication networks, GPS (or transponders), and fixed wayside signal devices to send and receive data about the location, direction, and speed of trains, as a generic PTC architecture is shown in Figure 7. If train crew fails to properly operate within specified safety parameters (e.g. train moving above the speed limit), the PTC system would activate an audible warning in the locomotive first to alert the train engineer. If the engineer does not reduce the train's speed, the PTC system would automatically apply the train brakes and bring the train to a positive stop without the engineer's assistance.[38]

The Rail Safety Improvement Act of 2008 mandated the implementation of PTC by 31 December 2018, or alternatively 31 December 2020 under special conditions.[40–42] After the completion of a nationwide PTC network, some collisions, in particular of the ones caused by the failure to obey signals and violation of train speed rules, can largely be prevented. To our knowledge, there is no comprehensive database of PTC-preventable accidents available to the public. In this research, we estimate the rough-order magnitude of the collision risk that is potentially preventable by PTC, based on the primary accident causes and narratives in the FRA REA database. Specifically, if the primary cause and narrative of a specific collision indicate that this accident is within current PTC implementation territory and is part of PTC functioning cases, it may have been prevented by a compliant PTC system. Using this approach, from 2001 to 2015, over 300 freight train collisions involving 32 fatalities, 416 injuries, and over 200 million damage cost to infrastructure and rolling stock would have been prevented had PTC been installed and functioned. In particular, the top three collision causes, which are failure to obey or display signals (05H), violation of train speed rules (10H), and violation of mainline

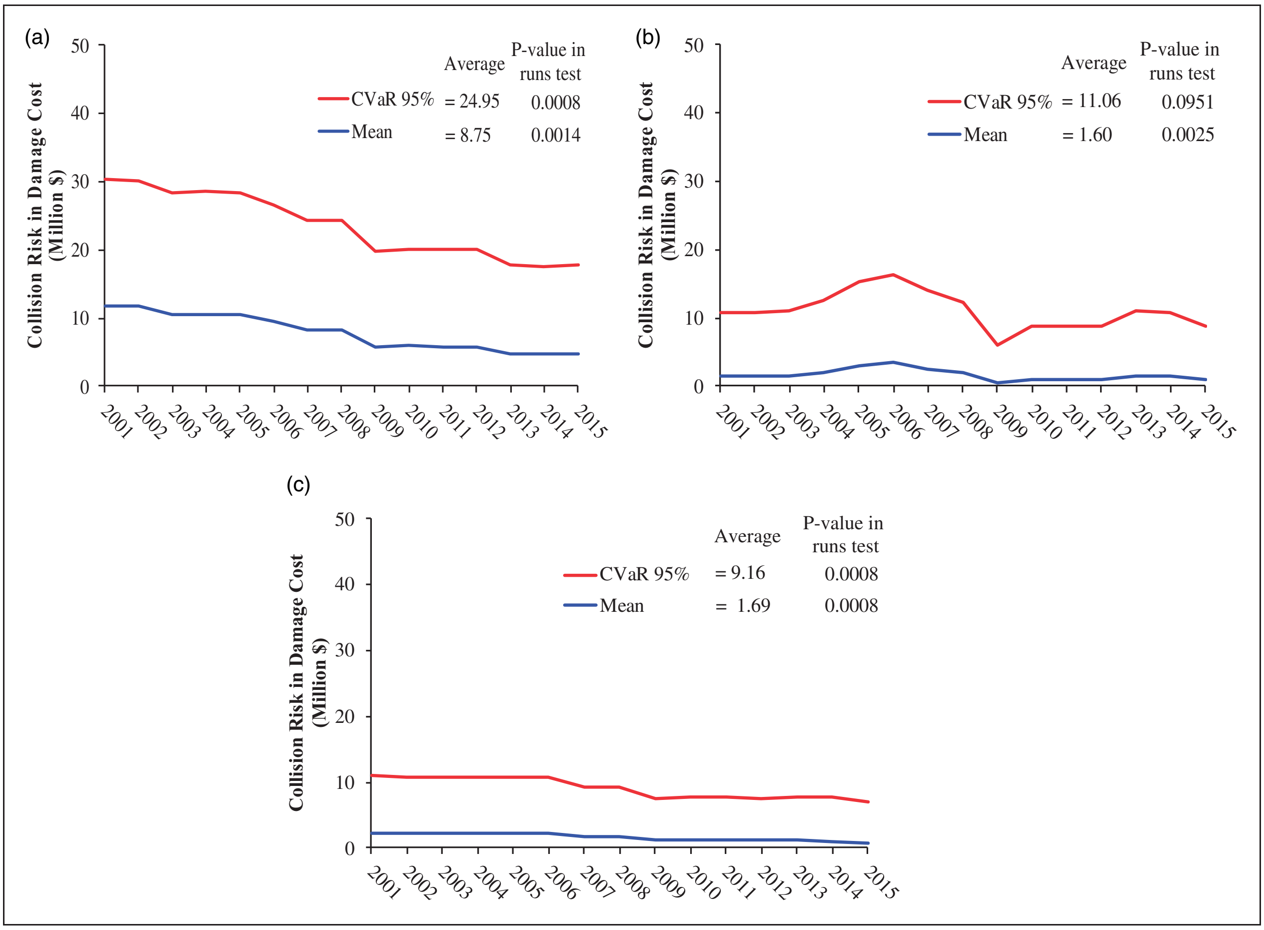

Figure 6. Freight train collision risk in terms of damage cost (million $) for major causes, 2001–2015. (a) Failure to obey or display signals (05H), (b) violation of train speed rules (10H), and (c) violation of mainline rules (08H). CVaR: conditional value at risk.

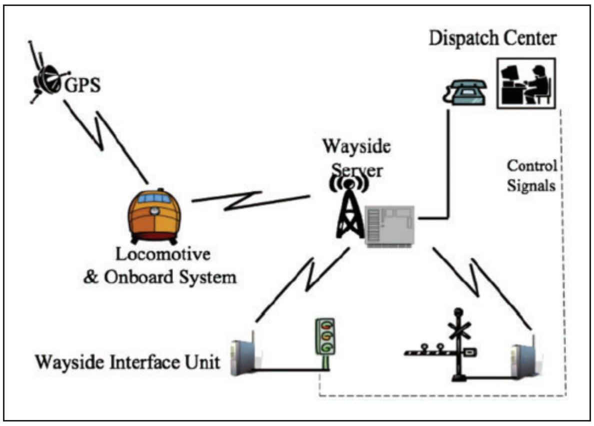

Figure 7. Basic PTC system architecture (cited from Hartong et al.[39]).

operating rules (08H), developed in "Causal analysis" section are mostly PTC preventable according to PTC basic functions and an FRA report[36] on human factors. Correspondingly, the estimated annual risk due to the three major causes presented in Table 8 could be reduced significantly once a nationwide PTC implementation is finished. Our current work is somewhat a "pre-PTC" safety analysis and can serve as a reference in support of developing future analysis to evaluate the safety effectiveness of PTC in terms of changing risk profiles, based on future accident data.

Conclusion

This paper develops a statistical risk analysis of freight train collisions in the United States based on the data from 2001 to 2015. Overall, collision rate per train-mile has an average annual declining rate of approximately 5%. There is no significant temporal trend of collision severity for either casualties or damage cost in the study period. Two alternative risk measures are used, including the expected consequence (mean) and CVaR. Compared to the expected consequence risk measure, the CVaR accounts for a small proportion of accidents with greater severities, and thus captures the low-probability–high-consequence characteristics of certain train accidents. Failure to obey or display signals, violation of mainline rules, and violation of train speed rules are the major collision causes, with the failure to obey or display signals having the greatest cause-specific risk due to its high frequency and high severity. Nationwide, the average freight train collision risk due to the most severe cause (failure to obey or display signals) is around \$9 million per year, or 14 casualties per year, using the expected consequence risk measure. When using the $CVaR_{95\%}$ (the average of worst 5% collisions), the risk due to the most severe cause is approximately \$25 million or 81 casualties.

Acknowledgements

The authors are solely responsible for all views, analyses, and errors in this paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors are partially funded by the Federal Railroad Administration of USDOT (DTFR5317C00008), at the time of writing this paper.

ORCID iD

Xiang Liu http://orcid.org/0000-0003-3126-8811

References

1. Muttram RI. Railway safety's safety risk model. *J Rail Rapid Transit* 2002; 216: 71–79.
2. Liu X. Analysis of collision risk for freight trains in the United States. *Transp Res Rec J Transp Res Board* 2016; 2546: 121–128.
3. Barkan CPL, Dick CT and Anderson RT. Railroad derailment factors affecting hazardous materials transportation risk. *Transp Res Rec J Transp Res Board* 2003; 1825: 64–74.
4. Liu X, Barkan CPL and Saat MR. Analysis of derailments by accident cause: evaluating railroad track upgrades to reduce transportation risk. *Transp Res Rec J Transp Res Board* 2011; 2261: 178–185.
5. Liu X, Saat MR and Barkan CPL. Analysis of causes of major train derailment and their effect on accident rates. *Transp Res Rec J Transp Res Board* 2012; 2289: 154–163.
6. Liu X, Saat MR and Barkan CPL. Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Anal Prev* 2013; 59: 87–93.
7. Saccomanno FF, Park PY-J and Fu L. Estimating countermeasure effects for reducing collisions at highway–railway grade crossings. *Accident Anal Prev* 2007; 39: 406–416.
8. Chadwick SG, Saat MR and Barkan CPL. Analysis of factors affecting train derailments at highway–rail grade crossings. In: *Presented at 91st annual meeting of the Transportation Research Board*, 2012.
9. FRA Rail Equipment Accident (6180.54) database, https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx (accessed 1 November 2017).
10. 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.* 2nd rev ed. Cambridge, MA: ADL, 1996.
11. Liu X, Saat MR and Barkan CPL. Probability analysis of multiple-tank-car release incidents in railway hazardous materials transportation. *J Hazard Mater* 2014; 276: 442–451.
12. Liu X, Liu C and Hong YL. Analysis of multiple tank car releases in train accidents. *Accident Anal Prev* 2017; 107: 164–172.
13. Liu X and Hong Y. Analysis of railroad tank car releases using a generalized binomial model. *Accident Anal Prev* 2015; 84: 20–26.
14. Evans AW. Estimating transport fatality risk from past accident data. *Accident Anal Prev* 2003; 35: 459–472.
15. Liu X. Statistical temporal analysis of freight-train derailment rates in the United States: 2000 to 2012. *Transp Res Rec J Transp Res Board* 2015; 2476: 119–125.
16. Miaou SP. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Anal Prev* 1994; 26: 471–482.
17. Hauer E. Overdispersion in modeling accidents on road sections and in empirical Bayes estimation. *Accident Anal Prev* 2001; 33: 799–808.
18. Wood GR. Generalised linear accident models and goodness of fit testing. *Accident Anal Prev* 2002; 34: 417–427.
19. Lord D, Washington SP and Ivan JN. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Anal Prev* 2005; 37: 35–46.
20. Lord D. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Anal Prev* 2006; 38: 751–766.
21. Oh J, Washington SP and Nam D. Accident prediction model for railway-highway interfaces. *Accident Anal Prev* 2006; 38: 346–356.
22. Mitra S and Washington S. On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Anal Prev* 2007; 39: 459–568.
23. Evans AW. Rail safety and rail privatization in Britain. *Accident Anal Prev* 2007; 39: 510–523.
24. Evans AW. Fatal train accidents on Europe's railways: 1980–2009. *Accident Anal Prev* 2011; 43: 391–401.
25. Hilbe JM. *Negative binomial regression*. Cambridge: Cambridge University Press, 2007.
26. Liu X. Risk analysis of transporting crude oil by rail: methodology and decision support system. *Transp Res Rec J Transp Res Board* 2016; 2547: 57–65.
27. Lin CY and Saat MR. Semi-quantitative risk assessment of adjacent track accidents on shared-use rail corridors. In: *Proceedings of the 2014 joint rail conferences*, Colorado Springs, CO, 2014.
28. Liu X. Optimizing rail defect inspection frequency to reduce the risk of hazardous materials transportation by rail. *J Loss Prev Process Ind* 2017; 48: 151–161.
29. Liu X. Risk comparison of transporting hazardous materials in unit trains versus mixed trains. *Transp Res Rec J Transp Res Board* 2017; 2608: 134–142.

30. Evans AW and Verlander NQ. What is wrong with criterion FN-lines for judging the tolerability of risk? *Risk Anal* 1997; 17: 157–168.
31. Spada M, Paraschiv F and Burgherr P. A comparison of risk measures for accidents in the energy sector and their implications on decision-making strategies. *Energy* 2018; 154: 277–288.
32. Soleimani H, Seyyed-Esfahani M and Kannan \G. Incorporating risk measures in closed-loop supply chain network design. *Int J Prod Res* 2014; 52: 1843–1867.
33. Sarkaylin S, Serraino G and Uryasev S. Value-at-risk vs. conditional value-at-risk in risk management and optimization. *Tutorials Oper Res* 2008; 52: 270–293.
34. Rockafellar RT and Uryasev S. Optimization of conditional value-at-risk. *J Risk* 2000; 2: 21–42.
35. Inui K, Kijima M and Kitano A. VaR is subject to a significant positive bias. *Stat Prob Lett* 2005; 72: 299–311.
36. Wreathall J, Roth E, Bley D et al. *Human factors considerations in the evaluation of processor-based signal and train control systems: human factors in railroad operations*. No. DOT/FRA/ORD-07/07. Washington DC: Federal Railroad Administration, 2007.
37. Federal Railroad Administration. *49 Code of Federal Regulations (CFR) 236 – Rules, Standards, and Instructions Governing the Installation, Inspection, Maintenance, and Repair of Signal and Train Control Systems, Devices, and Appliances*. 33 Federal Register 19684. Washington, DC: Federal Railroad Administration 2011.
38. Association of American Railroads (AAR). *Positive train control*. Washington, DC: Association of American Railroad, 2017.
39. Hartong M, Goel R and Wijesekera D. Positive train control (PTC) failure modes. *J King Saud Univ Sci* 2011; 23: 311–321.
40. Congress of the United States of America. *Rail Safety Improvement Act*. Public Law 110–432. Washington, DC: Congress of the United States of America, 2008.
41. Congress of the United States of America. *Positive Train Control Enforcement and Implementation Act of 2015 (PTCEI Act)*. H.R. 3651. Washington, DC: Congress of the United States of America, 2015.
42. Federal Railroad Administration (FRA). Positive train control, https://www.fra.dot.gov/ptc (2017, accessed 1 December 2017).

Appendix 1. Major collision cause by group

Cause group	Description	FRA cause codes	Code description
05H	Failure to obey or display signals	H201	Blue signal, absence of
		H202	Blue signal, imperfectly displayed
		H205	Flagging, improper or failure to flag
		H206	Flagging signal, failure to comply
		H207	Hand signal, failure to comply
		H208	Hand signal improper
		H209	Hand signal, failure to give/receive
		H217	Failure to observe hand signals given during a wayside inspection of moving train
		H218	Failure to comply with failed equipment detector warning or with applicable train inspection rules
		H219	Fixed signal (other than automatic block or interlocking signal), improperly displayed
		H220	Fixed signal (other than automatic block or interlocking signal), failure to comply
		H221	Automatic block or interlocking signal displaying a stop indication—failure to comply
		H222	Automatic block or interlocking signal displaying other than a stop indication—failure to comply
		H299	Other signal causes (detailed description in narrative)
10H	Train speed	H601	Coupling speed excessive
		H602	Switching movement, excessive speed
		H603	Train on main track inside yard limits, excessive speed
		H604	Train outside yard limits, in block signal or interlocking territory, excessive speed
		H605	Failure to comply with restricted speed in connection with the restrictive indication of a block or interlocking signal

(continued)

Continued

Cause group	Description	FRA cause codes	Code description
		H606	Train outside yard limits in nonblock territory, excessive speed
		H607	Failure to comply with restricted speed or its equivalent not in connection with a block or interlocking signal
		H699	Speed, other (detailed description in narrative)
08H	Mainline rules	H401	Failure to stop train in clear
		H402	Motor car or on-track equipment rules, failure to comply
		H403	Movement of engine(s) or car(s) without authority (railroad employee)
		H404	Train order, track warrant, track bulletin, or timetable authority, failure to comply
		H406	Train orders, track warrants, direct traffic control, track bulletins, written, error in preparation, transmission or delivery
		H499	Other main track authority causes (detailed description in narrative)