

Research Article

A Simultaneous Safety Analysis of Crash Frequency and Severity for Highway-Rail Grade Crossings: The Competing Risks Method

Amin Keramati,¹ Pan Lu ,² Xiaoyi Zhou,¹ and Denver Tolliver¹

¹Upper Great Plains Transportation Institute, North Dakota State University, NDSU Dept., 2880 P. O. Box 6050, Fargo, ND 58108-6050, USA

²Department of Transportation, Logistics and Finance, Upper Great Plains Transportation Institute, North Dakota State University, NDSU Dept., 2880 P. O. Box 6050, Fargo, ND 58108-6050, USA

Correspondence should be addressed to Pan Lu; pan.lu@ndsu.edu

Received 19 March 2020; Revised 19 June 2020; Accepted 16 July 2020; Published 3 August 2020

Academic Editor: Lu Gao

Copyright © 2020 Amin Keramati et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a mathematical model, the competing risks method, to investigate highway-rail grade crossing (HRGC) crash frequency and crash severity simultaneously over a 30-year period. The proposed competing risks model is a special type of survival analysis to accommodate the competing nature of multiple outcomes from the same event of interest; in this case, the competing multiple outcomes are crash severities, while event of interest is crash occurrence. Knowledge-gain-based benefits to be discovered through the application of this model and 30-year dataset are as follows: (1) a straightforward and integrated one-step estimation process that considers both crash frequency and severity likelihood in the same model, so direct hazard ranking considering both crash frequency and severity likelihood is possible; (2) interpretative effects of identified covariates from both the direction and magnitude perspectives; and (3) the long-term cumulative effect of contributors with the cumulative incidence function.

1. Introduction

The highway-rail grade crossing (HRGC) is a unique spatial location where two transportation modes intersect with each other at grade level. Crashes at HRGCs are attributed to potential points of conflict between roadway traffic and train traffic. The crashes usually have relatively more severe results because of the huge mass difference between vehicle and train. Moreover, the economic consequences of crashes at HRGCs can be further extended by traffic delays on both the railway and the roadway. Crash modeling has been crucially important and intensively researched by decision-makers and transportation agencies for decades. Stakeholders and decision-makers rely on those models to provide direction to improve HRGCs safety performance.

A large volume of literature has been found to analyze and predict transportation accidents. Most of the previous

research studies have focused on roadway intersection or roadway crashes [1–14]. Relatively little research effort has focused on HRGC accidents compared to roadway accidents [15–24]. Moreover, among all the previous HRGC accident analyses, the majority of them focus only either on crash frequency, often based on crossing inventory databases [19, 25–39], or on crash severity analysis, often based on historical crash police report databases [16, 40–52]. To understand and predict crash frequency and severity simultaneously and consistently is important for agencies seeking to improve safety so they can account for the common factors affecting both crash frequency and severity. Separate forecasting models help determine what factors affect the likelihoods of a crash occurrence or crash severity levels; however, there are several application obstacles: (1) Identified contributors are not consistent. Policy-reported surface conditions are often used in severity models but are

not available for crash occurrence models. (2) Estimated crash severity likelihoods are conditional probabilities given crash occurring based on identified unique set of contributors and not transferable for agencies to calculate absolute probability for a specific crash level. For example, separate forecasting models could provide 20% crash likelihoods with one set of contributors, say *A* to *E*, and 25% level-one crash severity, 30% level-two crash severity, and 45% level-three severity with another set of contributors, say *D* to *H*. Because of *F*, *G*, and *H* contributors, the probabilities are not transferable among different models; however, safety improvement agencies need consistent and commonly available information to assist in safety improvement decision-makings based on both crash occurrence and crash severity. The same forecasting model to account for both crash frequency and severities with commonly available contributors is needed so that unmeasurable variance can be accounted for in the same error term and the estimated likelihoods can be directly used by agencies. Little research has tried to incorporate crash frequency and crash severity in the same model.

This paper identified and proposed a modeling approach, the competing risks method, for HRGC accident analysis. This method has been verified and applied extensively in the medical field. However, it has not been used in safety analysis. The authors investigate the model's interpretive capabilities in frequency of crashes and crash severity simultaneously through the application of highway-rail crash and inventory data.

2. Literature on Crash Frequency and Severity

Previous HRGC safety studies focus either on crash frequency or on crash severity. Due to the random, discrete, and nonnegative nature of crash data, generalized linear models (GLMs) have been the most common technique for modeling crash frequency. Heydari and Fu [28] investigated a Poisson–Weibull model. Other statistical models [19, 33, 36, 53] such as zero inflated, hurdle, and generalized event count models were adopted to address data issues such as the excess number of zero accidents and overdispersions/underdispersions. Previous literatures [51, 54–56] also used data mining methods such as the hierarchical tree-based regression technique [39], neural network [23, 57], and random forests [58] to predict train-vehicle crash frequencies or severities at highway-rail grade crossings. Recently, Heydari et al. [27] presented a method to compare different geographic areas in terms of prespecified safety performance, which can measure crash frequency of a given type.

There is a vast body of previous researches focusing on modeling crash severity outcomes. Most of the modeling techniques are discrete choice approaches because of the discrete nature of crash severity levels. Hao and Daniel [45] and Abdel-Aty and Keller [57] applied ordered probit model in the United States considering crash severities that are naturally ordered. Hao and Daniel [44] followed up on the previous work and adopted an ordered probit model by considering various control measures at HRGCs. Eluru et al.

[40] present a latent ordered response model considering ten years of crash data at HRGCs in the United States. Hu et al. [46] applied a generalized logit model to assess factors affecting crash severity at Taiwan's railroad grade crossings. Recently, Zhao et al. [51] used binary logit models and a generalized linear mixed model to study the association of potential factors with pedestrian injury severity levels using ten years of data at HRGCs in the United States.

All the abovementioned methods are very useful to model and understand crash frequency and severity outcomes separately. However, to predict crash frequency and the likelihood of crash severity outcomes simultaneously is important for agencies that need to account for the common factors affecting crash frequencies and severities based on the same set of available data. Unaccounted variables that affect crash occurrence and severity will be accounted for in error terms for each separate model. However, it is likely that these error terms are interrelated because they are describing the same concerns. Abdel-Aty and Nawathe [59] proposed a two-step approach to estimate crash frequency based on simulated geometric and traffic exposure information and then estimate crash severity with a neural network. Zalinger et al. [60] try to develop an integrated hazard regression model that considered both the number of crashes and their severity. However, in the final model development, the number of crashes and their severity are treated as accident history and only the number of accidents is selected as the hazard. Ye et al. [61] developed a simultaneous modeling approach for crash analysis, but the model only considers crash frequencies by collision types simultaneously using crash data rather than inventory data.

This paper attempts to further contribute to the HRGC safety by presenting a modeling methodology that decision-makers can use to estimate crash frequency and crash severity outcomes simultaneously. Moreover, the research will investigate the contributors' effects on the rate of crash and severity outcomes and their long-term cumulative effects over time.

3. Modeling Methodology

The competing risks model is a special type of survival analysis designed to correctly estimate the marginal probability of incidence outcomes when more than one cause of failure is possible. This method is intensively used in medical research [62–68] to study patient deaths attributable to competing events such as cardiovascular and non-cardiovascular causes. Survival analysis intends to estimate the probability of an occurrence of an event of interest. In transportation safety analysis, the event of interest is crashes. A unique feature of survival data is that not all targets experience the failure (e.g., crash) by the end of the analysis period. This feature is true for HRGCs. This phenomenon is referred to as censoring. Furthermore, transportation crash analysis is suitable for the competing risks model if each HRGC is viewed as a patient with a crash indicating a survival failure and where different severity outcomes at each HRGC are seen as different causes of survival failure or outcomes. In this study, crash severity outcomes can be

property damage only (PDO), injury, and fatality and they can be considered as competing outcomes for crashes. Figure 1 indicates the modeling structure with the three severity levels. The model starts with an initial state of “no crash” at year zero, which is 1989 for all records in this study. So all 3,310 crossings included in the study have zero crash record at the year 1989. As time passes, some crossings experience crashes. As shown in Figure 1, there are a total of 261 PDO triggered failures, 147 injury triggered failures, and 67 fatal triggered failures for a total of 3,310 crossings within the 30-year analysis period (1990 to 2018). Note, crossings with multiple crashes are excluded from the study database.

Consequently, each record in the dataset has three main variables of status (D), time (t), and crash severity (j), whereby competing risk algorithm is able to estimate the crash occurrence and severity likelihood. Status (D) is a binary variable that is equal to 1 if crash occurs and 0 otherwise. Considering t as the occurrence crash time, model is able to use variables D and t to estimate the crash occurrence likelihood. Crash severity variable can be equal to 1, 2, or 3 indicating crash severity levels of PDO, injury, and fatality, respectively. Therefore, if D related to one of the crash severity levels j is equal to 1 for a specific record at time t , the model will be able to estimate probability of crash occurrence ($D = 1$) with severity level j by time t . Note, in the situation where a crossing status associated with all severity levels is equal to 0 ($D = 0$) for whole study time span ($t \in [0, 30]$), crossing is counted as censored. Details about censored data and the way the competing model deals with them are explained in the Modeling Methodology section.

3.1. Competing Risks Model and Cause-Specific Hazard Function. The observable data in competing risks models can be represented by (1) the time of failure T or time of crash occurrence, (2) the cause of failure of D or crash severity levels (PDO, injury, and fatality), and (3) a covariate vector Z indicating contributors such as AADT and percentage of trucks. Equation (1) indicates the *cause-specific hazard function*, which is the fundamental concept in competing risks models. The cause-specific hazard function describes the instantaneous rate of K^{th} event failure in subjects. It is event-free currently:

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}. \quad (1)$$

Competing risks model is a multivariate failure time model, where each individual patient is assumed to have a potential failure time for each type of failure [24]. The earliest of these failures is actually observed and the others are latent. If \tilde{T}_k is considered as the time to failure of cause k , only $T = \min\{\tilde{T}_k\}$ and D can be observed, while D is an index variable specifying which event happened first. In competing risks models, if the endpoint of interest has not yet occurred at the end of the observation window, the event time is right censored. Therefore, in this study, the crossings that have no crash records during the 30-year study period are considered as right censored data. One of the advantages of the competing risks model is the ability

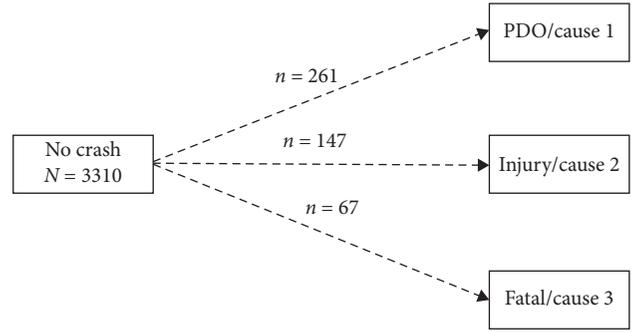


FIGURE 1: HRGC crash competing risks modeling structure.

to consolidate and utilize all the available data, even for the crossings that have no crash records. Previous studies only considered crossings with crash records [40, 48, 69]. Therefore, if some individuals (crossing records in HRGC safety analysis) are censored for all events by end of study, they have $D = 0$, and an extra censoring distribution $C \sim G$ is introduced, which is assumed to be independent of all the other events [70].

3.2. Regression on Cause-Specific Hazards. The cause-specific hazard of cause K , $\lambda_k(t | Z)$, for a crossing record with covariate values $Z = (Z_1, \dots, Z_p)$ is estimated in the following equation [70]:

$$\lambda_k(t | Z) = \lambda_{k,0}(t) \exp(\beta_k^T Z), \quad (2)$$

where the hazard ratio of crash is the instantaneous risk of failure from a particular severity level k for subject with covariate vector Z which is defined as $\exp(\beta_k^T Z)$ in equation (2). It is the conditional probability that a crossing with contributor vector Z has crash with severity level k at time t given it is crash-free just before time t . In equation (2), $\lambda_{k,0}(t)$ is the baseline cause-specific hazard of severity level k , and β_k indicates the contributors' effects on severity level k . The baseline hazard of severity level k can be estimated by the Breslow estimator, which is described by the following equation [71]:

$$\Delta \tilde{A}_k(t, \hat{\beta}) = \frac{\Delta N_k(t)}{S_k^{(0)}(\hat{\beta}, t)}, \quad (3)$$

where $\hat{\beta}$ is the maximum likelihood estimator of β and $\Delta N_k(t)$ denotes the number of records having crash with severity level k at t , and $S_k^{(0)}(\hat{\beta}, t)$ indicates the number of records at risk. The term $S_k^{(0)}$ represents the weighted risk set and is estimated in

$$S_k^{(0)}(\beta, t) = \sum_{i=1}^n \exp(\beta^T Z_{ki}(t)) Y_{ki}(t), \quad (4)$$

where $Y_{ki}(t)$ denotes the at-risk process; that is, $Y_{ki}(t) = 1$, and if the record i is at risk for severity level k at time t , the time point is just before time t . Note that, to estimate the effect of covariates on crash frequency, $\Delta N_k(t)$ in equation (3) is defined as the number of records that have crashes with any severity level.

The cause-specific function assumes independent censoring in estimating HR and coefficients. For instance, when a HRGC crash resulting in PDO is the event of interest, crossings with crashes resulting in other severities such as injury and fatality will be treated as censored observations. In other words, a crossing coded as PDO crash failure at time t is no longer at risk of injury crash or fatal crash failure at time t and will be treated as a censored observation. It is not easy to test whether a crossing that failed (crash occurred) in PDO would have failed from injury if the crossing did not fail from PDO, since the possible injury failure is unobservable for the crossings that actually are a PDO failure. But, in safety applications, dependence between competing risks should exist. In other words, a crossing having PDO crash can also experience a crash from other severity levels. Accordingly, the cumulative incidence function allows for estimation of the probability of the occurrence of an event while taking a competing risk into account, which does not assume the independence of competing events.

3.3. Cumulative Incidence Function. Cause-specific densities have the property of summing to the overall density as $\lambda(t) = \sum_k \lambda_k(t)$. Thus, the integral of the cause-specific density $I_k(t)$ is the cumulative incidence function (CIF) of crash severity level k . It describes the probability of crash occurrence with severity level k by time t , which is expressed in

$$I_k = \int_0^t \lambda_k(t) S(t) dt = \text{prob}(T \leq t_j, D = k), \quad (5)$$

where the overall survival probability $S(t)$ estimates the overall probability of failure not happening at time t . Equation (6) denotes the estimation of $S(t)$ at t without considering the crash severities that are estimated by the Kaplan–Meier estimator [70]:

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right). \quad (6)$$

In equation (6), let $0 < t_1 < t_2 < \dots < t_N$ be the ordered distinct crash time. If d_{kj} denotes the number of records having a crash with severity level k at t_j , then $d_j = \sum_{k=1}^K d_{kj}$ estimates the total number of crashes that occurred at t_j . The number of records at risk, n_j , indicates the number of records that are in follow-up status and have not had a crash until time t_j . A discretized version of cause-specific hazard of equation (1) can be revised as

$$\lambda_k(t_j) = \text{Prob}(T = t_j, D = k | T > t_{j-1}). \quad (7)$$

$\lambda_k(t_j)$ can be estimated by $\widehat{\lambda}_k(t_j) = d_{kj}/n_j$ indicating the proportion of records at the risk of crash occurrence with severity level k . To simplify and to estimate the effect of covariates on cumulative probability, equation (6) can be expressed as equation (8) considering the crash severity level k :

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \sum_{k=1}^k \widehat{\lambda}_k(t_j)\right). \quad (8)$$

The estimator of the cumulative incidence function (CIF) is expressed in

$$\widehat{I}_k(t) = \prod_{j:t_j \leq t} \widehat{S}(t_j) \left(\frac{d_{kj}}{n_j}\right). \quad (9)$$

4. Case Study

4.1. Data. Three main data resources are used for this research: (1) North Dakota (ND) roadway network, railway network, roadway intersections, and HRGCs from North Dakota Department of Transportation; (2) highway-rail grade crossing accident/incident data from the Federal Railway Administration (FRA); and (3) the highway-rail grade crossing inventory from FRA. The data include all reported crashes/incidents and their associated information, current and historical crossing inventory information for each crossing, and geometric features relative to the connecting highway and railways during the analysis period in North Dakota.

The final database contains information for 3,310 crossings, including 475 crash records and 2,835 no-crash records (total 3,310 records) for ND public HRGCs from 1989 to 2018 with three crash severity levels: PDO, injury, and fatality. Table 1 summarizes all the variables used in the study. Most records experienced no crash (86%) and the proportions of crash severities with PDO, injury, and fatality are 8%, 4%, and 2%, respectively. The key variables inputs are determined based on data availability and intuitive judgement, and the model will provide key variables with their significance based on model's statistical significance test. More detailed variables information associated with all crossings with or without crash records over 30-year period can be found in Table 1. In this study, not only is the crash information updated through 30-year analysis period for each crossing but also all contributors' inventory values for 30 years are kept to account for changes, so range, mode, and mean values of a variable are provided in Table 1.

4.2. Estimated Coefficient and Hazard Ratio Interpretation. Table 2 represents estimated contributors' coefficient (Coe) and hazard ratio (HR) for each severity level (PDO, injury, and fatality) and crash frequency. The estimated coefficient (β_k) and hazard ratio ($\exp(\beta_k^T Z)$) are estimated based on equations (2) and (3). For a categorical variable, HR indicates the relative risk for a crossing with that contributor's value level compared to the reference level. For a continuous contributor, HR is interpreted as the relative independent risk associated with a one-unit change in variable [73]. The regression coefficient from a cause-specific hazard model indicates the magnitude of the corresponding change in the cause-specific hazard function associated with a one-unit change in the contributor, while HR indicates the magnitude of the corresponding change in crash likelihood. As indicated in Table 2, positive Coe of 0.6 and HR of 1.82 indicate an 82% increase in PDO crash likelihood associated with crossings with passenger train service compared to freight train service. Negative Coe of -0.8 and HR of 0.45 indicate a

TABLE 1: Descriptive statistics of key variables.

Variable	Values	Frequency	Mean or percentage	Min	Max	Mode
Crash information	No crash	2835	86%			
	PDO	261	8%			
	Injury	147	4%			
	Fatality	67	2%			
Type of train service	Freight	2855	86%			
	Intercity passenger	455	14%			
Train detection	None	2450	74%			
	Constant warning time	438	13%			
	Motion detection	42	1%			
	DC	379	11%			
Is commercial power available?	Yes	2202	67%			
	No	1108	33%			
Is roadway/pathway paved?	Yes	625	19%			
	No	2685	81%			
Total daylight through trains			7.26	0	35	5
Total night time through trains			6.42	0	33	4.75
Total switching trains			0.77	0	12	0.55
Maximum train speed			41.02	5	79	37
Annual average daily traffic (AADT)			360.16	5	23438	45
Percent trucks			14.06	1	22.67	15.44
Distance to nearby roadway intersection			199.28	0.775	2502	27.04
Smallest crossing angle			68.37	7.869	90	90
Number of traffic lanes			1.812	1	4	2

Note. One important influential variable, warning device type, is not included in the model mainly due to unreliable data quality: (1) historical warning device types for each crossing for 30 years are not readily available and (2) inventory data contains some quality issues and results of warning device type could not be cross-validated. This could be the reason why not many long-term studies exist in the literature and some counterintuitive countermeasure effectiveness results in the literature [72]. Moreover, the focus of the study is to demonstrate the model's capability to model crash and severity likelihoods and its interpretive capability to provide contributors' long-term and instantaneous effects; contributors can be easily included in the model when they become available.

55% decrease in PDO crash likelihood associated with crossings intersected with an unpaved highway compared to a paved highway. The numerical value of HR can be any positive number with an HR of 1, indicating lack of association (likelihood of change is no different than zero), an HR greater than 1, suggesting an increased risk, and an HR less than 1, indicating a reduced risk.

According to Table 2, in general, intercity passenger train service (compared to freight train service), constant warning time (CWT) train detection system (compared to no train detection system), daytime train traffic, train speed, roadway traffic, and number of roadway lanes have positive impacts on crash likelihood while nighttime train traffic, direct current (DC) train detection system (compared to no train detection system), no commercial power available (compared to commercial power being available), and truck percentage have negative crash frequency impacts. Traffic exposure factors, such as daytime train traffic, roadway traffic, and train speed, all have positive impact on the likelihood of HRGC crashes, which meet expectations and agree with previous research results. However, nighttime traffic's negative impact on the likelihood of HRGC crashes is an interesting finding in the research. The potential

rationales need further research investigation. This negative impact may be caused by the operating changes. More nighttime idling trains switching to nighttime operating trains would reduce the daytime train traffic and reduce the conflict probability at grade crossings since, in general, roadway traffic is concentrated during the daytime.

One can see from Table 2 that some factors have significant impact on certain crash severity likelihoods but not on others. These results can be underestimated because of the independent censoring assumption. However, according to equation (5), CIF is estimated without the assumption of competing risks independency. Consequently, calculated competing events marginal likelihood indicates the competing events' dependency and is more accurate. Table 1 indicates that one factor has significant impact on all three severity likelihoods: night train traffic. Night train traffic has a negative impact on PDO and injury crashes but has a positive impact on fatality crashes. The potential rationale for the positive impact on fatality crashes and negative impact on PDO and injury crashes could be because nighttime drivers are less aware of the environment and existence of a HRGC because of reduced visibilities. In addition, nighttime drivers are more likely to operate their

TABLE 2: Estimated coefficient and hazard ratio.

Variable	PDO		Injury		Fatality		Crash	
	Coef	HR (CI: 95%)	Coef	HR (CI: 95%)	Coef	HR (CI: 95%)	Coef	HR (CI: 95%)
Type of train service (reference: freight)								
Intercity passenger	0.6**	1.82 (1, 3)	-0.2	0.82 (0.4, 2)	0.7	2.01 (0.8, 5)	0.4**	1.50 (1, 2)
Train detection (reference: none)								
CWT	0.2	1.22 (0.7, 2)	0.5**	1.65 (1, 3)	1***	2.72 (1, 8)	0.4***	1.50 (1.2, 2)
DC	-0.7***	0.50 (0.3, 0.8)	-0.6	0.55 (0.2, 1)	-2.0	0.14 (0.03, 2)	-0.7***	0.5 (0.3, 0.7)
Is commercial power available? (reference: yes)								
No	-0.1	0.90 (0.6, 1)	-0.7***	0.50 (0.3, 0.8)	0.3	1.35 (0.80, 2)	-0.2*	0.80 (0.6, 1)
Is roadway/pathway paved? (reference: yes)								
No	-0.8***	0.44 (0.3, 0.7)	-0.4	0.67 (0.4, 1)	-0.2	0.82 (0.4, 2)	-0.6***	0.50 (0.4, 0.7)
Total daytime through trains	0.2***	1.22 (1, 1)	0.1	1.11 (1, 1)	-0.4	0.67 (0.4, 1)	0.2***	1.22 (1.1, 1.3)
Total nighttime through trains	-0.2***	0.82 (0.7, 0.9)	-0.2*	0.82 (0.7, 1)	0.5*	1.65 (0.9, 3)	-0.1***	0.90 (0.8, 1)
Total switching trains	0.01	1.01 (0.8, 1)	0.03	1.03 (0.8, 1)	0.5**	1.65 (1, 3)	0.03	1.03 (0.9, 1.2)
Maximum train speed	0.004	1 (1, 1)	0.05**	1.05 (1, 1)	0.03	1.03 (1, 1)	0.02***	1 (1, 1)
Annual average daily traffic	0.00009**	1 (1, 1)	0.00004	1 (1, 1)	0.0001	1 (1, 1)	0.00008***	1 (1, 1)
Percent trucks	-0.08**	0.92 (0.9, 1)	-0.1***	0.90 (0.8, 1)	0.04	1.04 (0.9, 1)	-0.09***	0.91 (0.9, 1)
Distance to nearby intersecting roadway	-0.001**	1 (1, 1)	0.0002	1 (1, 1)	0.0005	1 (1, 1)	-0.0003	1 (1, 1)
Smallest crossing angle	-0.002	1 (1, 1)	-0.01**	1 (1, 1)	0.006	1 (1, 1)	-0.004	1 (1, 1)
Number of traffic lanes	0.4*	1.50 (1, 2)	0.1	1.11 (0.7, 2)	-0.01	1 (0.4, 2)	0.3**	1.35 (1.1, 1.6)

*Significant at a 90% confidence level. **Significant at a 95% confidence level. ***Significant at a 99% confidence level.

vehicles at higher speeds; thus more severe crashes could result from the increase in nighttime train traffic [40], making it less likely that nighttime crashes will be PDO or injury-only crashes.

It is found that, with a one-unit increase in train speed, crash likelihood increased by 5% for injury crashes and by 3% for fatality crashes but the likelihood of a change in PDO and overall crashes is not significantly different than zero.

Another interesting finding from the research is that the distance between the crossings to the nearest intersection has negative impact on the likelihood of PDO crashes. HRGCs with longer road distances to the nearest intersections are less likely to have PDO crashes, which is expected. This is because the longer distance provides a larger vehicle storage capability, which reduces the potential impact of roadway intersection operations and reduces potential PDO crashes. However, this longer distance has no significant impact on more severe crashes such as injury or fatality crashes.

Table 3 summarizes the importance ranking information based on the cause-specific HR. The instantaneous risk changes associated with HR and contributors are defined as “% impact,” and the contributors are ranked based on this value. Table 3 indicates that “train service” has the highest impact on PDO, and “train detection” has the highest impact on injury crash, fatality crash, and crash occurrence likelihood.

As indicated earlier, the hazard ratio reflects critical information regarding the contributor’s influence on

instantaneous crash and severity probabilities. However, the significance of the contributors can be underestimated because of the independent censoring assumption. For example, the effect for one risk such as a PDO crash may reflect the influence of competing risks such as in injury crash or a fatality crash. To truly understand contributors’ effects on hazard ratio and long-term crash probabilities when considering competing characteristics as in this case, a crossing that failed by PDO could have potentially failed by injury or fatal crash too. To capture the unobserved dependence among failure types, the cumulative-incidence-based effect analysis should be conducted.

4.3. Cumulative Probability over Time. Hazard ratio is a direct isolated influential indicator to a specific failure event. The isolated influential effects do not consider the same contributor’s effects on other competing failure events. In other words, they tend to underestimate the contributor’s impact when analyzing the marginal probability for cause-specific events considering the competing nature of multiple causes to the same event. Such analysis results can be very sensitive to different modeling techniques and provide very different contributing effects [74]. One of the advantages that the competing risks model can provide is to evaluate contributors’ long-term robust influence. Previous research [75] verified and indicated that a covariate has no effect on a

TABLE 3: Ranking variables based on hazard ratio.

Variable	PDO		Injury		Fatality		Crash	
	Rank	% impact	Rank	% impact	Rank	% impact	Rank	% impact
Type of train service:								
Intercity passenger	1	82%	5	18%	2	101%	2	49%
Train detection:								
CWT	5	22%	1	65%	1	172%	2	49%
DC	3	50%	3	45%	3	86%	1	50%
Is commercial power available?								
No	8	10%	2	50%	6	35%	6	18%
Is roadway/pathway paved?								
No	2	55%	4	33%	8	18%	3	45%
Total daytime through trains	6	22%	7	11%	7	33%	5	22%
Total nighttime through trains	7	18%	6	18%	4	65%	7	10%
Total switching trains	10	1%	11	3%	5	65%	9	3%
Maximum train speed	11	0.4%	10	5%	10	3%	10	2%
Annual average daily traffic	14	0.01%	14	0.004%	14	0.01%	13	0.01%
Percent trucks	9	8%	9	10%	9	4%	8	9%
Distance to nearby intersecting roadway	13	0.1%	13	0.02%	13	0.05%	12	0.03%
Smallest crossing angle	12	0.2%	12	1%	12	1%	11	0.4%
Number of traffic lanes	4	49%	8	11%	11	1%	4	35%

Note. % impact = $|HR - 1| \times 100$.

competing-event failure risk based on the cause-specific hazard analysis, but it still can indicate a significant impact on the cumulative risk probability of the competing event. A contributor, which has no direct influence on one type of failure event, may still be significantly associated with the cumulative probability (incidence) of that failure event, if that contributor has influences on a competing failure event. The estimated marginal probability of a certain failure event can be calculated with its cause-specific probability and the overall cumulative survival probability. Estimating cumulative probability of the failure events as the cumulative incidence function with equation (9) depends on HR for both the event of interest (crash occurrence) and the competing events (PDO, injury, and fatality crashes).

In the following study, “train service” and “train detection” are selected to perform the cumulative probability analysis due to the limited journal paper space and the facts that they are ranked as the top impact factors for PDO, injury, fatality crashes and crash occurrence likelihood. Moreover, the two-sample *t*-test method is used to quantify the significance of contributors’ effects on cumulative probability for crash occurrence and each severity level.

Predicted cumulative crash probabilities of each crash severity are estimated base on equations (8) and (9). To compare the change in predicted cumulative probability by changing train service types, two subsamples are created: one sample with freight train service only and the other with passenger train service only. The rest of the contributors’ values are controlled at a fixed level, mode value. Figure 2 indicates the 30-year predictions of cumulative crash severity and crash occurrence probabilities for “train service” in parts *a*, *b*, *c*, and *d*, respectively. Figure 3 describes the cumulative probability analysis results for train detection systems.

From Figure 2, one can see the following: (1) Cumulative crash probabilities are increasing over time at different rates

under various conditions. (2) In general, crossings with intercity freight train service have higher increasing rate of cumulative probabilities for all risks except for injury compared to the crossings with freight train services. (3) The increase in cumulative crash probabilities is faster for PDO crash occurrence than for injury and fatality crashes for both types of train service. (4) The increasing rate differs significantly between the two types of train service for all crash risks except for injury. (5) According to part *c*, even the absolute magnitude of the increasing rate is relatively small between the two services but the proportion of increased fatal probability is almost doubled for crossings with passenger train services compared to crossings with freight train services.

From Figure 3, one can see that crossings with DC systems have reduced crash likelihood on all crashes and all severities, but crossings with CWT systems have higher values compared to those with no detection systems. The absolute probability differences are all very small and around 0.1%. The fatal crash probability is more than doubled for crossings with CWT detection system compared to crossings with no detection system, 224% over 30 years. Note that crossings with a DC train detection system installed are 79% less likely in cumulative fatal probability compared to the crossings with no automatic detection system.

The results for CWT may sound counterintuitive. Detailed safety countermeasure effectiveness with regard to train detection systems should be conducted in the future with actual before- and after-installation data. Train detection warning systems are designed to warn crossing users of an approaching train through certain automatic train detection methods. For the DC system, current flows from a battery through a fixed rail segment to the coil of a relay. The locations of the battery and relay define the location of the warning-trigger rail segment.

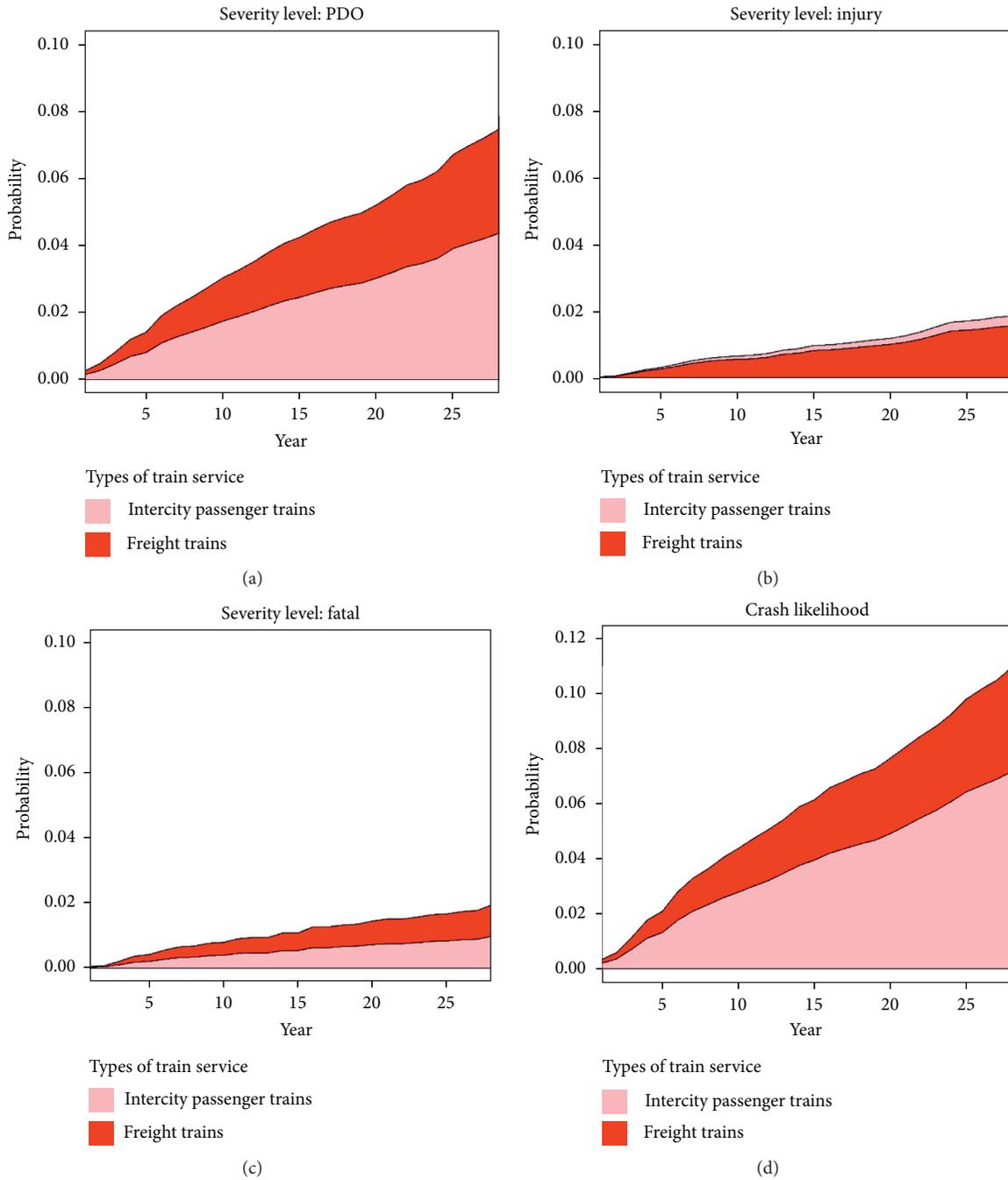


FIGURE 2: Cumulative probabilities of crash severity and crash occurrence for train service.

The DC method uses the rail as an energy conductor. When a train enters the segment, the axles short/shunt the circuit, which activates the crossing warning system to warn users of an approaching train. This method generates a warning based on the track occupation status, which has a fixed predefined distance from the crossing, usually between 1500 and 2000 feet. The CWT system is a smart system that is capable of determining the speed and location of an approaching train. Thus, it is able to predict when the train will arrive at the crossing. With the CWT system, a warning signal is activated intentionally to

provide a constant preselected warning time, usually 25 seconds. So, for a slow-moving train, the distance between the train and the crossing could be much closer than that for a faster-moving train. However, a CWT system is not able to measure a change in speed accurately, which results in variability in the actual warning time. For example, if a CWT system predicts the warning-activation time for a slow train and then the approaching train accelerates towards to the crossing, it will result in a less-than-desired warning time. This could be why crossings with CWT systems tend to have more crash results.

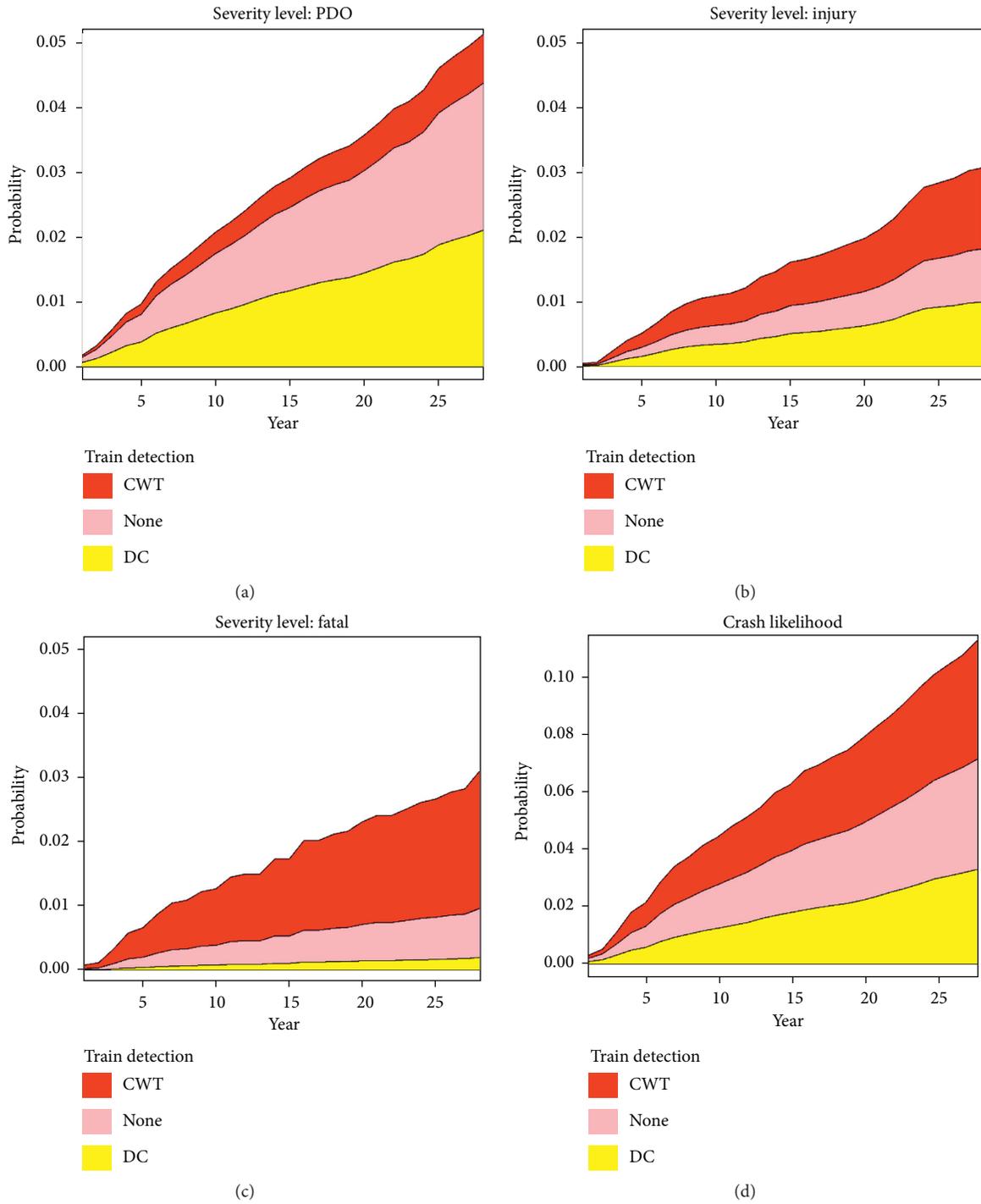


FIGURE 3: Cumulative probabilities of crash severity and crash occurrence for train detection.

The average annual probability increasing rates during the 30-year study period and the two-sample *t*-test results are summarized in Table 4. To apply the two-sample *t*-test, the average annual crash probability increasing rates of all records in the study dataset are estimated based on their cumulative incidence, and the test is applied for the selected categorical variable for each group level. One can see that, on average, PDO crash likelihood increased by about 0.26% each year for the crossings with passenger train service

compared to 0.152% for crossings with freight train services. On average, PDO crash likelihood at crossings with passenger train service increases by about 0.108% annually, which is 71% increase compared to freight train service category. Based on the *t*-test, train service change significantly impacts PDO crash likelihood at a 99% significance level.

Compared to Table 2, one can also see that crossings with intercity passenger service were identified as not significant

TABLE 4: Annual average crash probability increasing percentages for train service and detection system.

Variables	PDO % and change		Injury % and change		Fatality % and change		Crash % and change	
<i>Train service</i>								
Freight	0.152	NA	0.062	NA	0.033	NA	0.247	NA
Intercity passenger	0.26	0.108*** 71%	0.052***	-0.01 -16%	0.066***	0.033 100%	0.378***	0.131 53%
<i>Train detection</i>								
None	0.152	NA	0.062	NA	0.033	NA	0.247	NA
CWT	0.178***	0.026 17%	0.105***	0.043 69%	0.107***	0.074 224%	0.39***	0.143 58%
DC	0.073*	-0.079 -52%	0.034***	-0.028 -45%	0.007***	-0.026 -79%	0.114***	-0.133 -54%

to instantaneous injury crash likelihood compared to crossings with freight train services without considering competing risks. However, crossings with intercity passenger service were identified to significantly impact cumulative injury probability compared to freight train services when considering competing risks. The same change is observed for CWT systems for PDO crashes and for DC systems for injury and fatality crashes. As mentioned before, estimating coefficients and HRs through cause-specific hazard function are based on the assumption of independent censoring, which estimates coefficient and HR associated with each severity level separately. However, CIF is able to indicate whether a crossing is at the risk of one of crash severity levels (e.g., injury) and also would be at risk of the other crash severity levels (e.g., PDO or fatality), which means that CIF estimation is free of any events independence assumptions. Correspondingly, the significance difference between the two tables originates from the dependence of contributors' effects on competing events.

5. Conclusions and Discussion

The competing risks model is applied to examine crash frequency and severity simultaneously at public highway-rail grade crossings in North Dakota from 1989 to 2018. The competing risks model demonstrates its capability to identify specific contributors and to simultaneously model crash occurrence and crash severity likelihoods. It is able to produce easy-to-interpret results such as the estimated coefficients, hazard ratios, and cumulative probabilities. It also demonstrates its ability to capture the dependence of contributors' effects on competing events.

A few important findings are identified with the research:

- (1) Train services, train detection system, commercial power availability, roadway surface, train traffic, roadway traffic, train speed, truck percentage, and number of road lanes are identified as having significant impact on crash occurrence likelihood.
- (2) Crossings with passenger train services have higher PDO crash risk compared to crossings with freight train service, but they perform no differently in terms of instantaneous injury and fatal crash risk. Crossings with increased night train traffic are more likely to have fatality crashes and are less likely to have

PDO and injury crashes. Increasing the distance between the crossing and the nearest roadway intersection will decrease PDO crash risk but will have no impact on instantaneous injury or fatality crash risk. Increasing the smallest acute crossing angle may decrease injury crash risk. However, it will not impact instantaneous PDO and fatality crash risk.

- (3) Except for night train traffic, which is found to significantly impact all three crash severities, contributors are found to only have direct cause-specific impacts on certain crash severities but not on all of them.
- (4) Contributors, such as a CWT train detection system for PDO crashes, may not be significant on instantaneous risks, but they can be identified as significant contributors from a cumulative probability perspective when considering dependence competing risks.
- (5) Based on cumulative probabilities, the annual PDO crash probability increase is 0.108% for crossings with passenger train service compared to crossings with freight train service, which accounts for about 76% of the increase. Fatal probability increasing rate reduction is 0.033% for crossings with passenger train service compared to freight train service. However, this accounts for a 100% reduction. Crossings with DC detection systems tend to have lower crash increasing rates compared to crossings with no detection systems, but crossings with CWT detection systems tend to have higher crash increasing rates compared to those with no detection systems. Crossings with freight train service have about a 0.131% reduction in crash occurrence probability annual increasing rate compared to crossings with passenger train service.

The competing risks modeling demonstrates its capability to identify risk factors and the marginal probability for crash severity and crash occurrence simultaneously. The model is designed to identify and summarize the contributors and risk probabilities. However, the study does not suggest countermeasure effectiveness. A detailed countermeasure effectiveness analysis should be conducted with available supporting data.

Data Availability

The data that support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest associated with this publication.

Authors' Contributions

Research idea and study concept initiation were contributed by Dr. Pan Lu and Dr. Denver Tolliver. Study design was proposed by Dr. Pan Lu, Dr. Denver Tolliver, and Amin Keramati. Data collection was performed by Amin Keramati and Dr. Pan Lu. Model construction was carried out by Amin Keramati and Dr. Pan Lu. Lab/software resource support was provided by Dr. Denver Tolliver. Analysis and interpretations of results were performed by Amin Keramati, Xiaoyi Zhou, Dr. Pan Lu, and Dr. Denver Tolliver. Draft manuscript preparation was executed by Amin Keramati, Dr. Pan Lu, and Xiaoyi Zhou. All the authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

The authors express their gratitude to North Dakota State University and the Mountain-Plains Consortium (MPC), a University Transportation Center funded by the U.S. Department of Transportation, for their support.

References

- [1] Q. Cai, M. Abdel-Aty, and J. Lee, "Macro-level vulnerable road users crash analysis: a bayesian joint modeling approach of frequency and proportion," *Accident Analysis & Prevention*, vol. 107, pp. 11–19, 2017.
- [2] K. Geurts, I. Thomas, and G. Wets, "Understanding spatial concentrations of road accidents using frequent item sets," *Accident Analysis & Prevention*, vol. 37, no. 4, pp. 787–799, 2005.
- [3] Y. Hao, L. Xu, B. Qi, T. Wang, and W. Zhao, "A machine learning approach for highway intersection risk caused by harmful lane-changing behaviors," in *Proceedings of the CICTP 2019*, pp. 5623–5635, Nanjing, China, July 2019.
- [4] H. Huang, H. Zhou, J. Wang, F. Chang, and M. Ma, "A multivariate spatial model of crash frequency by transportation modes for urban intersections," *Analytic Methods in Accident Research*, vol. 14, pp. 10–21, 2017.
- [5] S. Islam and J. Brown, "A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama," *Accident Analysis & Prevention*, vol. 108, pp. 163–171, 2017.
- [6] S. Kumar, D. Toshniwal, and M. Parida, "A comparative analysis of heterogeneity in road accident data using data mining techniques," *Evolving Systems*, vol. 8, no. 2, pp. 147–155, 2017.
- [7] J. Lee, M. Abdel-Aty, and Q. Cai, "Intersection crash prediction modeling with macro-level data from various geographic units," *Accident Analysis & Prevention*, vol. 102, pp. 213–226, 2017.
- [8] L. Li, S. Shrestha, and G. Hu, "Analysis of road traffic fatal accidents using data mining techniques," in *Proceedings of the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 363–370, IEEE, London, UK, June 2017.
- [9] M. Paul, "Safety assessment at unsignalized intersections using post-encroachment time's threshold—a sustainable solution for developing countries," in *Lecture Notes in Civil Engineering*, pp. 117–131, Springer, Berlin, Germany, 2019.
- [10] X. Qin, J. N. Ivan, and N. Ravishanker, "Selecting exposure measures in crash rate prediction for two-lane highway segments," *Accident Analysis & Prevention*, vol. 36, no. 2, pp. 183–191, 2004.
- [11] M. B. Ulak, E. E. Ozguven, R. Moses et al., "Assessment of traffic performance measures and safety based on driver age and experience: a microsimulation based analysis for an unsignalized t-intersection," *Journal of Traffic and Transportation Engineering*, vol. 6, no. 5, 2019.
- [12] N. Veeramisti, A. Paz, M. Khadka, and C. Arteaga, "A clusterwise regression approach for the estimation of crash frequencies," *Journal of Transportation Safety & Security*, pp. 1–31, 2019.
- [13] X. Wang and M. Abdel-Aty, "Temporal and spatial analyses of rear-end crashes at signalized intersections," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1137–1150, 2006.
- [14] Z. Zheng, P. Lu, and B. Lantz, "Commercial truck crash injury severity analysis using gradient boosting data mining model," *Journal of Safety Research*, vol. 65, pp. 115–124, 2018.
- [15] H. Cho and L. R. Rilett, "Modeling signalized intersections near highway-railroad grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1973, no. 1, pp. 149–156, 2006.
- [16] H. Ghomi, M. Bagheri, L. Fu, and L. F. Miranda-Moreno, "Analyzing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: a comparative study," *Traffic Injury Prevention*, vol. 17, no. 8, pp. 833–841, 2016.
- [17] K. Haleem, "Investigating risk factors of traffic casualties at private highway-railroad grade crossings in the United States," *Accident Analysis & Prevention*, vol. 95, pp. 274–283, 2016.
- [18] A. Khattak, M. Gao, and Z. Luo, "The impact of removing centerline barriers from highway-rail grade crossings on drivers' violations," *Journal of Transportation Safety & Security*, vol. 4, no. 4, pp. 295–307, 2012.
- [19] P. Lu and D. Tolliver, "Accident prediction model for public highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 90, pp. 73–81, 2016.
- [20] L.-W. Tung and A. Khattak, "Distracted motor vehicle driving at highway-rail grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2476, no. 1, pp. 77–84, 2015.
- [21] H. Yue and E. Jones, "Archiving capability of spatio-temporal data in different highway railroad grade crossing (HRGC) databases," in *Proceedings of the Annual Intelligent Transportation System Conference*, Madeira, Portugal, September 2010.
- [22] S. Zhao and A. J. Khattak, "Factors associated with self-reported inattentive driving at highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 109, pp. 113–122, 2017.
- [23] Z. Zheng, P. Lu, and D. Pan, "Predicting highway-rail grade crossing collision risk by neural network systems," *Journal of*

- Transportation Engineering, Part A: Systems*, vol. 145, no. 8, Article ID 4019033, 2019.
- [24] Z. Zheng, P. Lu, and D. Tolliver, "Decision tree approach to accident prediction for highway-rail grade crossings: empirical analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2545, no. 1, pp. 115–122, 2016.
- [25] R. D. Austin and J. Carson, "An alternative accident prediction model for highway-rail interfaces," *Accident Analysis & Prevention*, vol. 34, no. 1, pp. 31–42, 2002.
- [26] R. Guadamuz-Flores and J. Aguero-Valverde, "Bayesian spatial models of crash frequency at highway-railway crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2608, no. 1, pp. 27–35, 2017.
- [27] S. Heydari, L. Fu, L. Thakali, and L. Joseph, "Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: applications to grade crossing safety analysis," *Analytic Methods in Accident Research*, vol. 19, pp. 33–48, 2018.
- [28] S. Heydari and L. Fu, "Developing safety performance functions for railway grade crossings: a case study of Canada," in *Proceedings of the 2015 Joint Rail Conference American Society of Mechanical Engineers (ASME)*, Article ID V001T06A017, San Jose, CA, USA, March 2015.
- [29] S.-R. Hu, C.-S. Li, and C.-K. Lee, "Model crash frequency at highway-railroad grade crossings using negative binomial regression," *Journal of the Chinese Institute of Engineers*, vol. 35, no. 7, pp. 841–852, 2012.
- [30] S. Hu and J. Lin, "Effect of time to train arrive on crash frequency at highway-railroad grade crossings: a general classification regression model," in *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2012.
- [31] A. Khattak, A. Sharma, and Z. Luo, "Implications of using annual average daily traffic in highway-rail grade crossing safety models," 2012.
- [32] A. Khattak and Z. Luo, "Pedestrian and bicyclist violations at highway-rail grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2250, no. 1, pp. 76–82, 2011.
- [33] J. Lee, D. Nam, and D. Moon, "A zero-inflated accident frequency model of highway-rail grade crossing," in *Proceedings of the Transportation Research Board Annual Meeting*, Washington, DC, USA, January 2004.
- [34] J. C. Medina and R. F. Benekohal, "Macroscopic models for accident prediction at railroad grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2476, no. 1, pp. 85–93, 2015.
- [35] H. Millegan, X. Yan, S. Richards, and L. Han, "Evaluation of effectiveness of stop-sign treatment at highway-railroad grade crossings," *Journal of Transportation Safety & Security*, vol. 1, no. 1, pp. 46–60, 2009.
- [36] J. Oh, S. P. Washington, and D. Nam, "Accident prediction model for railway-highway interfaces," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 346–356, 2006.
- [37] F. F. Saccomanno, P. Y.-J. Park, and L. Fu, "Estimating countermeasure effects for reducing collisions at highway-railway grade crossings," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 406–416, 2007.
- [38] F. F. Saccomanno and X. Lai, "A model for evaluating countermeasures at highway-railway grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1918, no. 1, pp. 18–25, 2005.
- [39] X. Yan, S. Richards, and X. Su, "Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 64–74, 2010.
- [40] N. Eluru, M. Bagheri, L. F. Miranda-Moreno, and L. Fu, "A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings," *Accident Analysis & Prevention*, vol. 47, pp. 119–127, 2012.
- [41] W. Fan, M. R. Kane, and E. Haile, "Analyzing severity of vehicle crashes at highway-rail grade crossings: multinomial logit modeling," *Journal of the Transportation Research Forum*, vol. 54, pp. 39–56, 2015.
- [42] K. Haleem and A. Gan, "Contributing factors of crash injury severity at public highway-railroad grade crossings in the U.S.," *Journal of Safety Research*, vol. 53, pp. 23–29, 2015.
- [43] W. Hao and J. Daniel, "Driver injury severity related to inclement weather at highway-rail grade crossings in the United States," *Traffic Injury Prevention*, vol. 17, no. 1, pp. 31–38, 2016.
- [44] W. Hao and J. Daniel, "Motor vehicle driver injury severity study under various traffic control at highway-rail grade crossings in the United States," *Journal of Safety Research*, vol. 51, pp. 41–48, 2014.
- [45] W. Hao and J. R. Daniel, "Severity of injuries to motor vehicle drivers at highway-rail grade crossings in the United States," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2384, no. 1, pp. 102–108, 2013.
- [46] S.-R. Hu, C.-S. Li, and C.-K. Lee, "Investigation of key factors for accident severity at railroad grade crossings by using a logit model," *Safety Science*, vol. 48, no. 2, pp. 186–194, 2010.
- [47] Y. Kang and A. Khattak, "Cluster-based approach to analyzing crash injury severity at highway-rail grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2608, no. 1, pp. 58–69, 2017.
- [48] J. Liu and A. J. Khattak, "Gate-violation behavior at highway-rail grade crossings and the consequences: using geo-spatial modeling integrated with path analysis," *Accident Analysis & Prevention*, vol. 109, pp. 99–112, 2017.
- [49] C. Ma, W. Hao, W. Xiang, and W. Yan, "The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents," *Journal of Advanced Transportation*, vol. 2018, Article ID 9841498, 10 pages, 2018.
- [50] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1666–1676, 2011.
- [51] S. Zhao, A. Iranitalab, and A. J. Khattak, "A clustering approach to injury severity in pedestrian-train crashes at highway-rail grade crossings," *Journal of Transportation Safety & Security*, vol. 11, no. 3, pp. 305–322, 2018.
- [52] S. Zhao and A. Khattak, "Motor vehicle drivers' injuries in train-motor vehicle crashes," *Accident Analysis & Prevention*, vol. 74, pp. 162–168, 2015.
- [53] Z. Ye, Y. Xu, and D. Lord, "Crash data modeling with a generalized estimator," *Accident Analysis & Prevention*, vol. 117, pp. 340–345, 2018.
- [54] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Analysis & Prevention*, vol. 108, pp. 27–36, 2017.
- [55] D. Lee, J. Warner, and C. Morgan, "Discovering crash severity factors of grade crossing with a machine learning approach," in *Proceedings of the Joint Rail Conference American Society of*

- Mechanical Engineers Digital Collection*, Salt Lake City, UT, USA, April 2019.
- [56] C. Yang, E. Trudel, and Y. Liu, "Machine learning-based methods for analyzing grade crossing safety," *Cluster Computing*, vol. 20, no. 2, pp. 1625–1635, 2017.
- [57] M. Abdel-Aty and J. Keller, "Exploring the overall and specific crash severity levels at signalized intersections," *Accident Analysis & Prevention*, vol. 37, no. 3, pp. 417–425, 2005.
- [58] X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, "Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree," *Reliability Engineering & System Safety*, vol. 200, Article ID 106931, 2020.
- [59] M. Abdel-Aty and P. Nawathe, "A novel approach for signalized intersection crash classification and prediction," *Advances in Transportation Studies*, vol. 9, 2006.
- [60] D. A. Zalinger, B. A. Rogers, and H. P. Johri, "Calculation of hazard indices for highway-railway crossings in Canada," *Accident Analysis & Prevention*, vol. 9, no. 4, pp. 257–273, 1977.
- [61] X. Ye, R. M. Pendyala, V. Shankar, and K. C. Konduri, "A simultaneous equations model of crash frequency by severity level for freeway sections," *Accident Analysis & Prevention*, vol. 57, pp. 140–149, 2013.
- [62] P. K. Andersen, L. S. Hansen, and N. Keiding, "Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process," *Scandinavian Journal of Statistics*, vol. 18, pp. 153–167, 1991.
- [63] M. Fiocco, H. Putter, C. J. H. Van De Velde, and J. C. Van Houwelingen, "Reduced rank proportional hazards model for competing risks: an application to a breast cancer trial," *Journal of Statistical Planning and Inference*, vol. 136, no. 5, pp. 1655–1668, 2006.
- [64] M. Fiocco, H. Putter, and J. C. Van Houwelingen, "Reduced rank proportional hazards model for competing risks," *Biostatistics*, vol. 6, no. 3, pp. 465–478, 2005.
- [65] R. B. Geskus, "On the inclusion of prevalent cases in hiv/aids natural history studies through a marker-based estimate of time since seroconversion," *Statistics in Medicine*, vol. 19, no. 13, pp. 1753–1769, 2000.
- [66] R. B. Geskus, F. A. Miedema, J. Goudsmit, P. Reiss, H. Schuitemaker, and R. A. Coutinho, "Prediction of residual time to aids and death based on markers and cofactors," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 32, no. 5, pp. 514–521, 2003.
- [67] T. A. Gooley, W. Leisenring, J. Crowley, and B. E. Storer, "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators," *Statistics in Medicine*, vol. 18, no. 6, pp. 695–706, 1999.
- [68] R. P. Van Rij, A.-M. De Roda Husman, M. Brouwer, J. Goudsmit, R. A. Coutinho, and H. Schuitemaker, "Role of ccr2 genotype in the clinical course of syncytium-inducing (si) or non-si human immunodeficiency virus type 1 infection and in the time to conversion to si virus variants," *The Journal of Infectious Diseases*, vol. 178, no. 6, pp. 1806–1811, 1998.
- [69] J. Liu, A. J. Khattak, S. H. Richards, and S. Nambisan, "What are the differences in driver injury outcomes at highway-rail grade crossings? untangling the role of pre-crash behaviors," *Accident Analysis & Prevention*, vol. 85, pp. 157–169, 2015.
- [70] H. Putter, M. Fiocco, and R. B. Geskus, "Tutorial in biostatistics: competing risks and multi-state models," *Statistics in Medicine*, vol. 26, no. 11, pp. 2389–2430, 2007.
- [71] L. C. De Wreede, M. Fiocco, and H. Putter, "The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 3, pp. 261–274, 2010.
- [72] E. H. Farr, "Summary of DOT Rail-Highway Crossing Resource Allocation Procedure-Revised," Federal Railroad Administration, Washington, DC, USA, 1987.
- [73] B. R. Logan, M.-J. Zhang, and J. P. Klein, "Regression models for hazard rates versus cumulative incidence probabilities in hematopoietic cell transplantation data," *Biology of Blood and Marrow Transplantation*, vol. 12, no. 1, pp. 107–112, 2006.
- [74] J. J. Dignam, Q. Zhang, and M. Kocherginsky, "The use and interpretation of competing risks regression models," *Clinical Cancer Research*, vol. 18, no. 8, pp. 2301–2308, 2012.
- [75] M. Wolbers, M. T. Koller, V. S. Stel et al., "Competing risks analyses: objectives and approaches," *European Heart Journal*, vol. 35, no. 42, pp. 2936–2941, 2014.