



Research Article

Quantification of Rear-End Crash Risk and Analysis of Its Influencing Factors Based on a New Surrogate Safety Measure

Qiangqiang Shangguan,^{1,2} Ting Fu^{ID, 1,2} Junhua Wang,^{1,2} Rui Jiang,³ and Shou'en Fang^{1,2}

¹The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

²College of Transportation Engineering, Tongji University, 4800 Cao'an Highway, Shanghai 201804, China

³Investment and Development Department, China Shandong International Economic & Technical Cooperation Group Ltd., 1822A Shandong Hi-speed Group Mansion, Jinan 250098, Shandong, China

Correspondence should be addressed to Ting Fu; ting.fu@mail.mcgill.ca

Received 18 February 2021; Accepted 22 April 2021; Published 30 April 2021

Academic Editor: Yanyong Guo

Copyright © 2021 Qiangqiang Shangguan 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.

Traditional surrogate measures of safety (SMoS) cannot fully consider the crash mechanism or fail to reflect the crash probability and crash severity at the same time. In addition, driving risks are constantly changing with driver's personal driving characteristics and environmental factors. Considering the heterogeneity of drivers, to study the impact of behavioral characteristics and environmental characteristics on the rear-end crash risk is essential to ensure driving safety. In this study, 16,905 car-following events were identified and extracted from Shanghai Naturalistic Driving Study (SH-NDS). A new SMoS, named rear-end crash risk index (RCRI), was then proposed to quantify rear-end crash risk. Based on this measure, a risk comparative analysis was conducted to investigate the impact of factors from different facets in terms of weather, temporal variables, and traffic conditions. Then, a mixed-effects linear regression model was applied to clarify the relationship between rear-end crash risk and its influencing factors. Results show that RCRI can reflect the dynamic changes of rear-end crash risk and can be applied to any car-following scenarios. The comparative analysis indicates that high traffic density, workdays, and morning peaks lead to higher risks. Moreover, results from the mixed-effects linear regression model suggest that driving characteristics, traffic density, day-of-week (workday vs. holiday), and time-of-day (peak hours vs. off-peak hours) had significant effects on driving risks. This study provides a new surrogate safety measure that can better identify rear-end crash risks in a more reliable way and can be applied to real-time crash risk prediction in driver assistance systems. In addition, the results of this study can be used to provide a theoretical basis for the formulation of traffic management strategies to improve driving safety.

1. Introduction

Statistics from the World Health Organization (WHO) show that the number of deaths caused by road traffic crashes is about 1.35 million each year, ranking eighth among all causes of death [1]. The serious consequences of traffic crashes have driven researchers to investigate the causes of the crashes. Among the many causes, driving behavior has been found to be a crucial one. For example, a study conducted by National Highway Traffic Safety Administration (NHTSA) found that driver-related factors account for 94% of the critical reasons of these crashes, and most studies indicated that traffic crashes can largely result from risky

driving behaviors [2–4]. To reduce casualties and mitigate injuries from traffic crashes, understanding and identifying the crash risk is essential.

Among different types of traffic crashes, rear-end crashes are recognized as one of the most common types [5, 6]. Statistics from the NHTSA show that the rear-end crashes accounting for 32.4% of all accident types that cause personal injury [7]. Since most rear-end crashes occurred in car-following situations, it has become crucial to identify the rear-end crash risk during car-following process and explore its influencing factors [8–10].

Despite the efforts on rear-end crash risk identification and analysis, several research gaps still exist. Measures,

including time-to-collision (TTC), stop distance index (SDI), deceleration rate to avoid crash (DRAC), and others, have been proposed to study driving risks [11–13]. However, TTC based on constant velocity assumption ignores the response of the following vehicle (FV) and the changes in the states of the vehicle pairs. Therefore, measures which take the mechanism of driver response and the development of the crash into account should be developed for better representing risks during car-following events. Besides, the other traditional surrogate measures of safety (SMoS) cannot fully reflect the crash probability and crash severity at the same time. In addition, driving risks change with driver's personal characteristics and environmental factors [3, 14–16]. In-depth research on the risk of rear-end crash risk and its influencing factors is essential to formulate effective countermeasures to reduce the risk of rear-end crash.

Judging from the previous studies, the traditional SMoS cannot fully consider the crash mechanism or fail to reflect the crash probability and crash severity at the same time. In addition, just few previous studies have been conducted based on large-scaled naturalistic driving data and comprehensively consider driver heterogeneity, behavioral characteristics, and environmental characteristics to study the influencing factors of rear-end crash risk from the perspective of microscopic car-following behavior.

Therefore, this study aims to propose a reliable measure to quantify the driving risk in the process of car-following and investigate the impact of various influencing factors on the rear-end crash risk considering driver's heterogeneity. For that purpose, a new rear-end crash risk index (RCRI) was introduced, which fully considers the crash mechanism and integrates crash probability and crash severity. A total of 16,905 car-following events were extracted from Shanghai Naturalistic Driving Study (SH-NDS). Crash risks under different influencing factors were analyzed and compared based on the proposed RCRI. Then, the mixed-effects linear regression model was then employed to study the impact of the behavioral characteristics and environmental factors on rear-end crash risk.

2. Literature Review

2.1. Overview of Studies on Surrogate Measures of Safety (SMoS). High-risk car-following behaviors, such as close-driving to the leading vehicle (LV), may lead to a high probability of an accident [17]. Based on temporal and spatial proximity, various rear-end crash risk indexes have been proposed that can be used to evaluate driving risks. Among time-based surrogate measures, time-to-crash (TTC) was widely used in practice [11]. Meanwhile, the risk of rear-end crash also depends on driver's crash avoidance behavior. The crash can also be avoided if the FV brakes in time. Herein, deceleration rate to avoid crash (DRAC) was then introduced to evaluate the braking requirement during vehicle conflicts to quantify the risk of rear-end crash [13]. Besides, maintaining a safety distance between the LV and the FV is the key to avoiding rear-end crashes. Therefore, the stop distance index (SDI), which is based on the concept of safe stopping distance of the FV, was proposed by Oh et al.

[12]. To mitigate the risk of rear-end crash, the FV should maintain the safe headway distance from the LV.

Although the traditional SMoS was widely used in quantifying the risk of rear-end crash, they also have some limitations. Kuang et al. [18] summarized three main limitations of the SMoS presented above: (i) the driver's response characteristics when experiencing a conflict are not considered in these SMoS; (ii) due to the requirements of the boundary conditions, any situation where the speed of the FV is lower than the speed of the LV is considered a safe situation, which may be unreasonable for the situation where the vehicle is traveling at a similar speed and the headway distance is small; (iii) the arbitrary selection of the index threshold will also make the result inaccurate.

Researchers have also proposed some new surrogate measures to address the abovementioned problems. Xie et al. [19] proposed the time-to-crash with disturbance (TTCD) to quantify the risk of rear-end crash. This indicator solves the problem of inaccurate risk identification when the speed of the FV is less than the speed of the LV. Besides, Shi et al. [20] derived a new hybrid indicator named key risk indicator (KRI), which integrates time integrated time-to-crash (TIT), crash potential index (CPI), and stopping sight distance (PSD). However, these proposed surrogate measures cannot make up all the above three limitations. In view of this, Kuang et al. [18] proposed a surrogate measure based on tree structure to evaluate the rear-end crash risk, namely, aggregated crash index (ACI), which considers many major factors in rear-end crash, including disturbance of LV, driver's reaction characteristics, and available braking capacity of the FV. However, this method is not convenient due to its complicated leaf nodes, which makes it complicated to apply. In addition, this measure only discusses the occurrence probability of a rear-end crash event without considering the severity of the crash. Therefore, it is crucial to propose a new SMoS to address the above limitations and better quantify the risk of rear-end crash.

2.2. Overview of Studies on Driving Risk-Related Factors. Driving risks are often considered to be related to multiple factors, including driver's individual characteristics and environmental factors [3, 14–16]. Several studies have explored the interaction between driving risks and various influencing factors. Table 1 summarizes these previous studies, which can be summarized into three main categories based on experimental methods: survey-based studies, driving simulation studies, and naturalistic driving studies.

Most of the initial research is based on reported survey data. However, the main problem is that the sample size of the data is limited and may be highly subjective. In addition, survey-based research pays more attention to the influence of driver characteristics on driving risks and ignores driving environmental factors. Alternatively, due to its high safety, controllability, and comprehensive data acquisition, driving simulators are used to study driver behavior characteristics to improve driving safety, especially in safety-critical conditions. However, studies based on driving simulation experiments are mostly limited to specific behavior in a limited

TABLE 1: Summary of studies on driving risk-related factors.

Author	Research purpose	Experimental methods	Sample size (number of participants)	Analytical methods	Main findings
Hayley et al. [23]	To investigate the influence of driver's personality characteristics on risky driving behavior.	Online survey	175 drivers (79 female) aged 18–64 years old.	Regression analyses	Risky driving has a greater correlation with emotion recognition and expression levels, and less correlation with age.
Harbeck et al. [24]	To examine risky driving in relation to psychological variables.	Online survey	601 participants (230 female) aged 17–25 years old.	Statistical analyses	The results of the proposed model show that young driver's risky driving is related to risk perception, response cost, and rewards.
Ulleberg and Rundmo [25]	To understand the underlying mechanism of risky driving behavior through the combination of personality traits and social cognitive methods.	Questionnaire	3942 participants (2208 female) aged 16–23 years old.	Structural equation model (SEM)	The personality of a driver indirectly affects risky driving behaviors mainly by influencing behavioral attitudes.
Tao et al. [2]	To investigate the relationship between gender, age, self-reported risky driving behaviors, and crash risk.	Questionnaire	511 participants (195 female) aged 36–45 years old.	Structural equation model (SEM)	The results show that both driving experience and dangerous driving behavior affect the risk of accidents. The driver's gender has little influence on dangerous driving behavior and accident risk.
Teye-Kwadjo [26]	To investigate the impact of driver's risk perception on risky driving behavior.	Questionnaire	519 participants (127 female) aged 20–59 years old.	Structural equation model (SEM)	The research results show that risk perception has an impact on drivers' risky driving behaviors. Male and female drivers and married and unmarried drivers have different preferences for risky driving behaviors.
Shangguan et al. [27]	To evaluate the rear-end crash risks under adverse environment.	Driving simulation	32 participants (12 female) aged 23–45 years old.	Survival analysis	The results show that the lower visibility leads to higher rear-end crash risk, and road alignment has a significant impact on crash risk.
Precht et al. [28]	To identify the main influencing factors contributing to driving risks.	Naturalistic driving data	108 trip segments.	Generalized linear mixed models (GLMMs)	Driving violations are related to anger, the presence of passengers, and personal differences. In addition, secondary tasks that cause distraction of the driver's visual attention and complex driving tasks are associated with high driving risks.
Pnina et al. [14]	To investigate the interactions between driving context and their associations with risky driving behaviors of young novice drivers.	Naturalistic driving data	81 teenager drivers (43 female), average age 16.48 years old ($SD = 0.33$).	Passion regression analyses	Driving own has a higher high-risk driving than shared vehicle, and driving during the day has a higher risky driving rate than driving at night.
Chen et al. [29]	To explore the contributing factors to crash risk during lane-changing process.	Naturalistic driving data	579 lane-changing vehicle groups	Mixed regression models	The distance between the lane-changing vehicle and the preceding vehicle in the lane before the lane-changing significantly affects lane-changing safety.

number of safety-critical scenarios and cannot truly reflect the external driving environment [21]. Research based on naturalistic driving data represents real-world driving situations. It is possible to extract significant driving behavior parameters from naturalistic driving data, such as speed, acceleration, relative position with surrounding vehicles, and environmental conditions to study the influencing factors of driving risk [22]. The collected data are more comprehensive and effective and may provide more valid results.

3. Data Preparation

3.1. Brief Introduction of the Shanghai Naturalistic Driving Study (SH-NDS). The real-world driving data used in this paper were collected by the SH-NDS, jointly conducted by Tongji University, General Motors (GM), and the Virginia Tech Transportation Institute (VTTI) from 2012 to 2016 [30, 31]. The 60 drivers participating in this naturalistic driving experiment are aged between 35 and 50 years old, and all of whom have a driving experience of more than five years. The total mileage that has been driven before participating in the experiment is more than 20,000 kilometers, and the average daily mileage is not less than 40 kilometers. Each driver drives the assigned experimental vehicle on the open road network, and the driving route is selected according to the driver's daily travel needs.

The video data of SH-NDS are mainly recorded by 4 cameras, which are installed in hidden locations that are not easy to observe. The video image is shown in Figure 1, which is composed of the front and rear vision of the vehicle, the driver's facial state, and the hand operation image.

3.2. Car-Following Event Extraction. In this study, the car-following events in the SH-NDS were extracted to analyze the influencing factors of rear-end crash risk. The SH-NDS data cover all daily trips of the drivers participated. Totally data of 18,242 trips were collected. In the SH-NDS, data were automatically collected using the data acquisition system, triggered by the ignition switch of the vehicle. Therefore, the database inevitably contains a large number of trip records that are not related to the research content (vehicle activities including fueling, car washing, maintenance, and other types of short-distance trip were all recorded). In addition, missing values and outliers also exist in the SH-NDS database. Data processing mainly includes the following four steps:

- (i) Step 1: eliminate invalid record files. Due to the large number of short-distance trip records in the SH-NDS database, considering the distance between adjacent entrances and exits on urban expressways, a trip includes at least entering the urban expressway, driving on the urban expressway, and exiting the urban expressway, so the trip files with the travel time less than 5 minutes should be removed.
- (ii) Step 2: remove driving data under discontinuous traffic flow. To extract the car-following event under

continuous traffic flow, a point-to-point map matching algorithm was used to match the driving trajectory measured by GPS with the electronic map road data to find the trip record on the urban expressway [32]. Then, each trip was verified through camera video recording to ensure the validity of the driving data selected for analysis. The weather and light conditions were also determined during the verification process.

- (iii) Step 3: handle missing values and outliers. In order to obtain data such as vehicle's speed, acceleration, relative space-time distribution of surrounding vehicles, and traffic environment conditions during the car-following process, data preprocessing is required. Linear interpolation was applied to deal with missing values. Then, outliers were eliminated based on pauta criterion, and the data were smoothed using the moving average filter.
- (iv) Step 4: extract data of car-following events. Car-following events were extracted by applying an automatic extraction algorithm proposed by Zhu et al. [33]; (1) radar target's identification number of $LV > 0$ and remained constant: guaranteeing the FV was preceded by the same LV; (2) $7 \text{ m} < \text{longitudinal distance between the FV and the LV} < 120 \text{ m}$: eliminating the congested-flow conditions; (3) lateral distance between the FV and the LV $< 2 \text{ m}$: ensuring that the FV and the LV were in the same lane; (4) duration of the car-following event $> 15 \text{ s}$: guaranteeing that each car-following event has enough data for analysis.

After the application of the above steps, 16,905 car-following events (about 135 h of total event duration) were extracted from 1,197 trips. Figure 2 shows the histogram of the duration of car-following events.

4. Methodology

The methodology of this research mainly includes three parts: (i) derivation of a new surrogate measure for rear-end crash risk, (ii) identification of influencing factors for rear-end crash risk, and (iii) mixed-effects linear regression for rear-end crash risk modeling and factor analysis.

4.1. Derivation of New Surrogate Measure for Rear-End Crash Risk

4.1.1. Rear-End Crash Mechanism. In the process of car-following maneuver, the driver will choose the appropriate speed and safe headway distance according to the movement status of the LV. The determination of the safe headway distance needs to consider the driver's reaction time and the vehicle's deceleration process. Otherwise, it may easily lead to a rear-end crash when the headway distance is too small. In this research, we imposed a hypothetical disturbance on the LV, assuming the LV decelerated at a certain deceleration rate. As can be seen in Figure 3, the FV will take appropriate evasive actions based on the initial driving

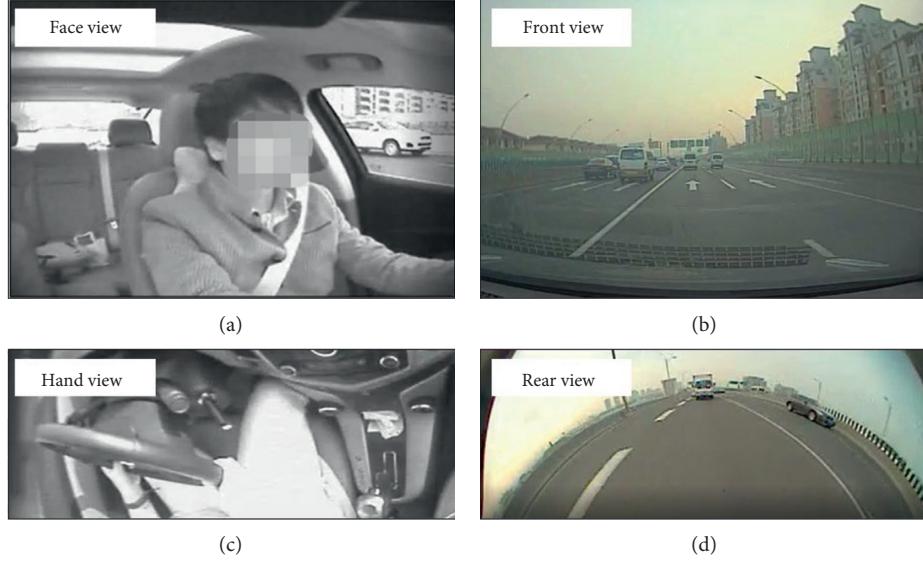


FIGURE 1: The video image of the SH-NDS.

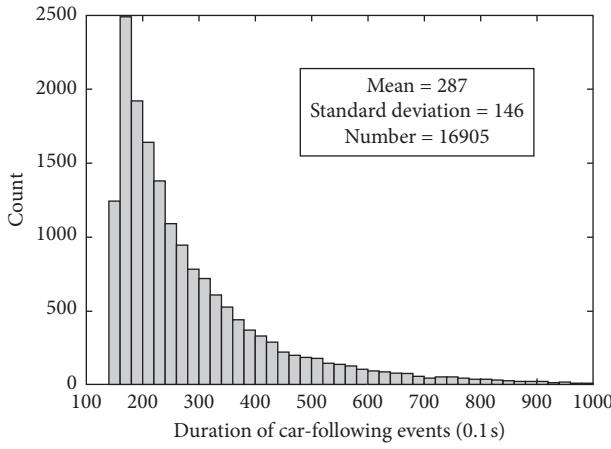


FIGURE 2: Histogram of the length of car-following events.

condition and the disturbance after reaction to avoid the crash. The crash outcome can be identified by evaluating the initial conditions, the disturbance, the driver's reaction characteristic, and the degree of evasive action [18, 34].

4.1.2. Rear-End Crash Risk Index (RCRI). Risk is the product of the possibility that a hazard event will occur and the consequence of the event. To address the limitations of SMoS mentioned above, we propose a new SMoS named RCRI, which considers the crash probability and crash severity at the same time, to quantify the risk of rear-end crash. According to the changes in the speed and distance of the LV and the FV before the crash, the process of rear-end crash can be divided into four categories, as shown in Figure 4 [35, 36].

Then, according to the characteristics of the rear-end crash, that is, the crash is plastic and the two vehicles tend to move together after the crash. In this study, the momentum

theorem was used to calculate the speed of the two vehicles after the rear-end crash, that is,

$$m_l v_{\text{pre_}l} + m_f v_{\text{pre_}f} = (m_l + m_f) v_c, \quad (1)$$

where m_l is the mass of the LV, m_f is the mass of the FV, $v_{\text{pre_}l}$ is the speed of the LV when the crash occurs, $v_{\text{pre_}f}$ is the speed of the FV when the crash occurs, and v_c is the speed of the LV and FV after the crash.

Calculate the energy loss ΔE of the two vehicles after rear-end crashes using the law of conservation of energy, as follows:

$$\begin{aligned} \Delta E &= \frac{1}{2} m_l v_{\text{pre_}l}^2 + \frac{1}{2} m_f v_{\text{pre_}f}^2 - \frac{1}{2} (m_l + m_f) v_c^2 \\ &= \frac{1}{2} \frac{m_l m_f}{(m_l + m_f)} (v_{\text{pre_}l} - v_{\text{pre_}f})^2. \end{aligned} \quad (2)$$

Therefore, this study uses the square of the absolute speed difference (SASD) at the time of the rear-end crash of two vehicles to express the severity of the rear-end crash.

To simplify the rear-end crash avoidance process, the model adopts two assumptions: (i) the braking process of the vehicle is regarded as a uniform deceleration process and (ii) the FV only adopts braking measures to avoid crash. Therefore, the braking stop distances of the two vehicles before the crash can be obtained, as shown in the following formula:

$$s_l = v_l t_c - \frac{1}{2} a_l t_c^2, \quad (3)$$

where s_l is the braking distance of the LV, v_l is the speed of the LV before braking, t_c is the time of crash, and a_l is the deceleration rate of the LV.

When $t_c \geq (v_l/a_l)$, the LV has been completely stopped, and the corresponding braking stop distance of LV is

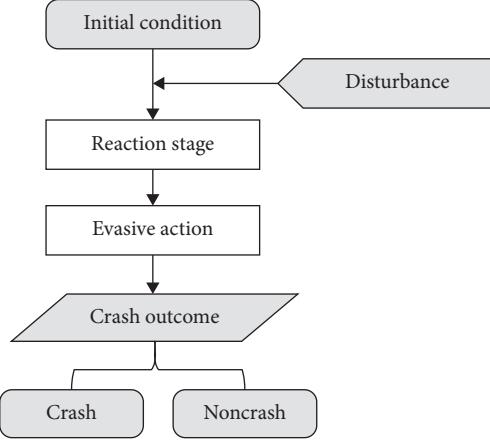


FIGURE 3: Schematic diagram of rear-end crash mechanism.

$$s_l = \frac{v_l^2}{2a_l}. \quad (4)$$

During the entire conflict process, during the reaction time t_R , the FV maintains a constant speed v_f and then decelerates at a constant deceleration rate a_f . When $0 \leq t_c \leq t_R$, the braking distance of the FV can be represented as

$$s_f = v_f t_c, \quad (5)$$

where s_f is the braking distance of the FV.

When $t_R \leq t_c \leq t_R + (v_f/a_f)$, the braking stop distance of the FV is

$$s_f = v_f t_R + v_f (t_c - t_R) - \frac{1}{2} a_f (t_c - t_R)^2. \quad (6)$$

Therefore, if the longitudinal distance between the FV and the LV is reduced to zero before the FV completely stopped, a crash will definitely occur, namely,

$$t_{c3} = \frac{(v_l - a_f t_R - v_f) \pm \sqrt{(a_f t_R + v_f - v_l)^2 + 2(a_l - a_f)((1/2)a_f t_R^2 + l)}}{a_l - a_f}. \quad (13)$$

If $a_l = a_f$, the solution is

$$t_{c3} = \frac{(1/2)a_f t_R^2 + l}{a_f t_R + v_f - v_l}, \quad (14)$$

$$\text{SASD}_3 = ((v_l - a_l t_{c3}) - (v_f - (t_{c3} - t_R) a_f))^2.$$

$$s_l + l = s_f, \quad (7)$$

where l is the initial gap between LV and FV.

For scenario 1, where $0 \leq t_{c1} \leq t_R$ and $0 \leq t_{c1} \leq (v_l/a_l)$, when the crash occurs,

$$v_l t_{c1} - \frac{1}{2} a_l t_{c1}^2 + l = v_f t_{c1}. \quad (8)$$

The solution is

$$t_{c1} = \frac{(v_l - v_f) \pm \sqrt{(v_f - v_l)^2 + 2a_l l}}{a_l}, \quad (9)$$

$$\text{SASD}_1 = ((v_l - a_l t_{c1}) - v_f)^2.$$

For scenario 2, where $(v_l/a_l) \leq t_{c2} \leq t_R$, when the crash occurs,

$$\frac{v_l^2}{2a_l} + l = v_f t_{c2}. \quad (10)$$

The solution is

$$t_{c2} = \frac{(v_l^2/2a_l) + l}{v_f}, \quad (11)$$

$$\text{SASD}_2 = (v_f)^2.$$

For scenario 3, where $0 \leq t_{c3} \leq (v_l/a_l)$ and $t_R \leq t_{c3} \leq t_R + (v_f/a_f)$, when the crash occurs,

$$v_l t_{c3} - \frac{1}{2} a_l t_{c3}^2 + l = v_f t_R + v_f (t_{c3} - t_R) - \frac{1}{2} a_f (t_{c3} - t_R)^2. \quad (12)$$

If $a_l \neq a_f$, the solution is

For scenario 4, where $t_c \geq (v_l/a_l)$ and $t_R \leq t_c \leq t_R + (v_f/a_f)$, when the crash occurs,

$$\frac{v_l^2}{2a_l} + l = v_f t_R + v_f (t_c - t_R) - \frac{1}{2} a_f (t_c - t_R)^2. \quad (15)$$

The solution is

$$t_{c4} = \frac{(v_f + a_f t_R) \pm \sqrt{(v_f + a_f t_R)^2 - 2a_f (v_l^2/2a_l) + l + (1/2)a_f t_R^2}}{a_f}, \quad (16)$$

$$\text{SASD}_4 = (v_f - (t_{c4} - t_R) a_f)^2.$$

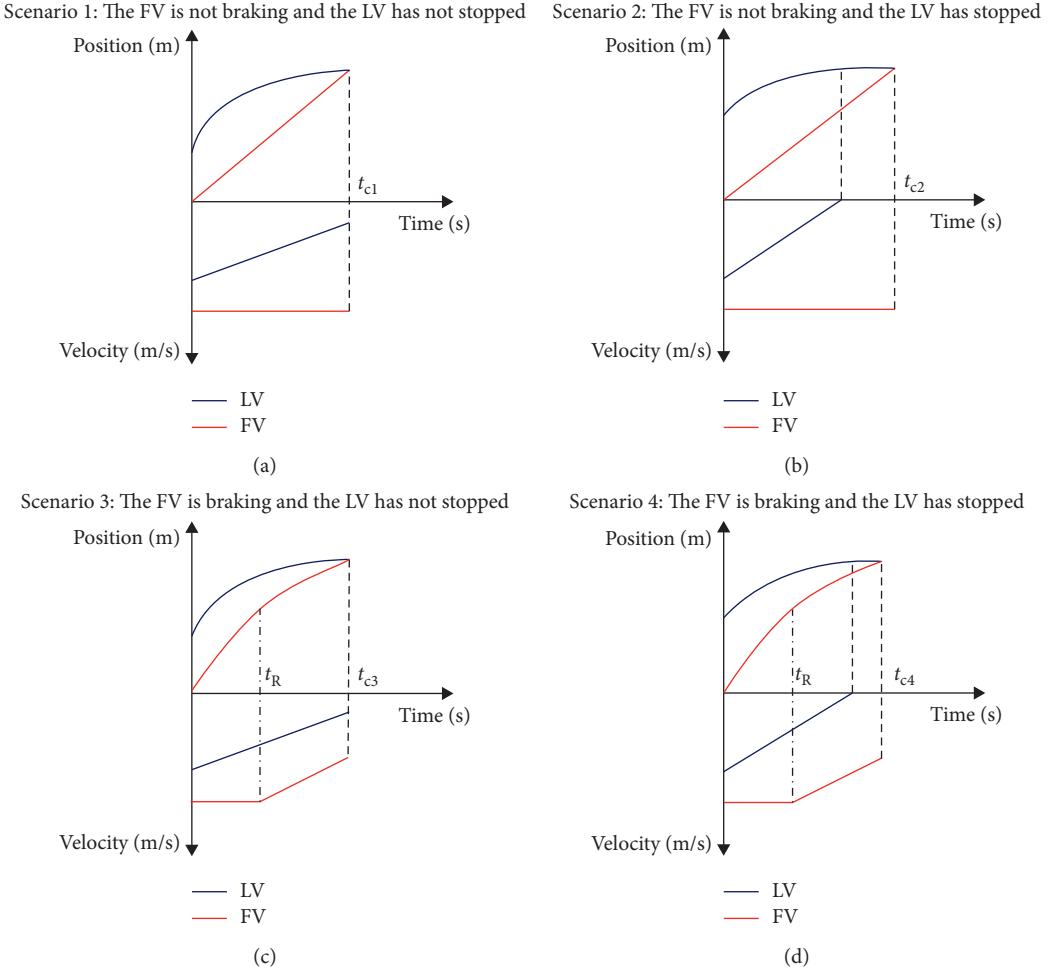


FIGURE 4: Descriptions of the four rear-end crash scenarios. Note. t_R – the reaction time of driver, t_c – the time of crash.

According to the previous studies, the deceleration rate taken by the LV follows a shifted gamma distribution (17.315, 0.128, 0.657), which was suggested and calibrated by Kuang et al. [18]. The reaction time of the FV follows a log-normal distribution (0.17, 0.44), and the braking coordination time is 0.175 s [37]. The maximum available deceleration rate (MADR) was assigned to be a truncated normal distribution with a mean of 8.45 m/s^2 and a variance of 1.4 m/s^2 between 4.23 m/s^2 and 12.68 m/s^2 [38].

The Monte Carlo simulation method was used to randomly select the deceleration rate taken by the LV, the reaction time of the driver, and the deceleration rate of the FV on the basis of the distribution function of each parameter mentioned above. According to the initial states of the LV and FV, the crash time (t) and the crash consequence (represented by SASD as discussed before) were calculated based on the above equations. Then, by integrating the possibility and consequences of the crash under the current car-following conditions, a new SMoS named RCRI can be calculated and expressed as below. It should be noted that the SASD needs to be normalized to a value between 0 and 1 before calculating RCRI. Therefore, the calculated RCRI range is also between 0 and 1:

$$\text{RCRI}_i = \frac{\sum_{j=1}^N \text{crash}_{ij} \times \text{SASD}_{ij}}{N}, \quad (17)$$

where RCRI_i represents the risk of rear-end crash at the i^{th} moment (0.1 s) in the car-following scenario; $N = 10,000$ is the number of random samples generated by the Monte Carlo simulation. When crash time t has a solution, that is, a crash occurs, $\text{crash}_{ij} = 1$; otherwise, $\text{crash}_{ij} = 0$.

It should be noted that SASD is dimensional, so it needs to be normalized before calculation.

4.2. Identification of the Influencing Factors for Rear-End Crash Risk

4.2.1. Definition of the Variables. As discussed in previous studies, the driving risk is mainly affected by driver's operational characteristics and the external driving environment [3, 14–16]. In order to quantify the various influencing factors for risks in the car-following process, this study extracted three categories of variables: behavioral variables, temporal variables, and environmental variables, as shown in Table 2.

TABLE 2: Summary of independent variables.

Categories	Variables	Conditions
Behavioral variables	Duration of car-following event	Continuous variable (0.1 s)
	Mean time headway	Continuous variable (s)
	Average speed difference	Continuous variable (km/h)
	Acceleration difference ratio	Continuous variable
Temporal variables	Day-of-week	Holiday, workday
	Time-of-day	Off peak, morning peak, evening peak
Environmental variables	Light condition	Daytime, nighttime
	Weather condition	Sunny, rainy
	Traffic density	High, median, low

The relevant variables of the driver's car-following behavior consider the duration of car-following event, time headway, and the driving speed and acceleration of the LV and FV. In order to eliminate the influence of the absolute value of the speed on the modeling, in terms of the speed indicator, this study adopted the average speed difference (ASD) between LV and FV, which is presented as

$$\text{ASD} = \frac{\sum_{t=1}^T |v_{l(t)} - v_{f(t)}|}{T}, \quad (18)$$

where $v_{l(t)}$ and $v_{f(t)}$ are the instantaneous speed of the LV and FV at its t -th record, and $t \in T$, T is the duration of car-following event.

Acceleration difference ratio (ADR) refers to the ratio of the standard deviation of the acceleration of the FV and LV. The ADR of one car-following event can be calculated as

$$\text{ADR} = \frac{\sigma_f}{\sigma_l}, \quad (19)$$

where σ_f and σ_l are the standard deviation of the acceleration of the FV and LV during one car-following period.

Peak hours increase the likelihood of congestion, resulting in a shortage of driving space, which in turn breeds risky driving behaviors. In order to consider the possible impact of the peak period, in this study, the car-following events within a day are divided into three categories: morning peak, evening peak, and off peak. Among them, the morning peak refers to 7:00 to 9:00, while the evening peak refers to 17:00 to 19:00. Besides, the traffic density was determined based on the speed of the FV and forward camera video recording, as described in Yang et al. [39].

To examine the effect of these independent variables on rear-end crash risk, the mean RCRI of a single car-following event was used as the dependent variable in this study.

4.2.2. Statistics Description of the Variables. As mentioned, we extracted 16,905 car-following events from 1,197 trips. Table 3 presents the descriptive statistics of the continuous variables and frequency information of the discrete independent variables in this study.

4.3. Mixed-Effects Linear Regression for Rear-End Crash Risk Modeling and Factor Analysis. In this research, all drivers participated in multiple car-following events, and car-following behavior characteristic variables were repeatedly collected from each participant. Therefore, the correlation problem between repeated observations will be exposed, that is, within-cluster correlation, and this problem can be solved by applying mixed-effects linear regression model [30, 40].

The mixed-effects linear regression model is an extension of the linear regression model, including fixed effects and random effects. Compared with the ordinary linear regression model, the mixed-effect linear regression model can well control the influence of the driver's personality factors. The results of the model can reflect the commonality of drivers and consider the internal correlation between samples, which is more suitable for solving the research problems.

In this paper, the fixed effects are the independent variables that the research focuses on. Besides, the drivers were treated as random effects to address the problem of within-cluster correlation.

Formally, the mixed-effects linear regression model can be written as

$$y = X\beta + Z\mu + \varepsilon, \quad (20)$$

where $y \in R^{N \times 1}$ is the dependent variable; $X \in R^{N \times p}$ is a matrix of the independent variables; $\beta \in R^{p \times 1}$ is the coefficients of the fixed effects; $Z \in R^{N \times q}$ is the matrix for random effects; $\mu \in R^{q \times 1}$ is the coefficients of the random effects; and $\varepsilon \in R^{N \times 1}$ is a column vector of the residuals. To recap,

$$\overbrace{y}^{N \times 1} = \underbrace{\overbrace{X}^{N \times p} \underbrace{\beta}_{p \times 1}}_{N \times 1} + \underbrace{\overbrace{Z}^{N \times q} \underbrace{\mu}_{q \times 1}}_{N \times 1} + \underbrace{\varepsilon}_{N \times 1}. \quad (21)$$

To better understand the structure of the model, here we provide an example where 16,905 (N) car-following events were collected from 58 (q) drivers. Our outcome y is the risk of car-following event. As mentioned above, we have 11 fixed effect predictors. The following equations represent the vectors and matrices provided in the previous equations:

TABLE 3: Summary of variables descriptive statistics.

<i>Independent variables (behavioral variables)</i>				
Variables	Mean	Std. Dev	Min	Max
Duration of car-following event (0.1 s)	286.66	146.44	151	1010
Mean time headway (s)	2.01	0.77	0.53	8.17
Average speed difference (m/s)	1.11	0.75	0.09	7.15
Acceleration difference ratio	0.91	0.38	0.00	4.08
<i>Independent variables (temporal variables)</i>				
Variables	Dummy variable		Count	
Day-of-week	Holiday		4200	
	Workday		12705	
	Off peak		9368	
Time-of-day	Morning peak		3873	
	Evening peak		3664	
<i>Independent variables (environmental variables)</i>				
Variables	Dummy variable		Count	
Light condition	Daytime		12456	
	Nighttime		4449	
Weather condition	Sunny		14587	
	Rainy		2318	
	High		1604	
Traffic density	Median		4495	
	Low		10806	
<i>Dependent variables (driving risk)</i>				
Variables	Mean	Std. Dev	Min	Max
RCRI	0.0031	0.0061	0.0000	0.7736

$$\begin{aligned}
 y &= \begin{bmatrix} \text{risk} \\ 0.013 \\ 0.002 \\ \dots \\ 0.006 \end{bmatrix}_{16,905 \times 1}, \\
 X &= \begin{bmatrix} \text{Intercept} & \dots & \text{Density} \\ 1 & \dots & 1 \\ 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & 2 \end{bmatrix}_{16,905 \times 11}, \\
 \beta &= \begin{bmatrix} 0.004 \\ 0.001 \\ 0.005 \\ \dots \\ -0.001 \end{bmatrix}_{11 \times 1}, \\
 Z &= \begin{bmatrix} \text{Driver}_1 & \dots & \text{Driver}_{58} \\ 1 & \dots & 0 \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \end{bmatrix}_{16,905 \times 58}.
 \end{aligned} \tag{22}$$

The random effects μ in the regression model are a column vector containing random intercepts. However, it is not necessary to estimate μ in actual regression modeling.

Instead, the model assumes that μ follows a normal distribution, with a mean of zero and a variance of σ^2 :

$$\mu \sim N(0, \sigma^2). \tag{23}$$

Parameters of all components were estimated using the mixed procedure in Stata/MP 16.0. The statistical significance level was set at 0.05.

5. Results and Discussion

5.1. Rear-End Crash Risk Identification Using SH-NDS Data and RCRI. To better illustrate the rear-end crash risk identification measure proposed in this study, TTC, DRAC, SDI, and RCRI were all employed to quantify the risk for one car-following event. According to the previous studies, the thresholds of each SMoS are chosen as follows. The threshold of TTC is normally chosen as 3 s [41], and the threshold of DRAC is 3.4 m/s^2 [42]. In the SDI calculation, it is assumed that the deceleration speed is 3.3 m/s^2 and the reaction time is 1.0 s [20]. A portion of the vehicle movement data during one car-following event is presented in Table 4. The risks identified at intervals of 0.1 s are shown in Figure 5.

As mentioned in the methodology, the RCRI is calculated based on the assumed disturbance. This measure can be used to quantify the risk in any scenario, even when the LV's speed is greater than that of the FV. Besides, the RCRI takes into account the most critical variables in crash mechanisms such as driver reaction characteristics and vehicle braking performance and comprehensively considers the crash probability and consequences. Moreover, the RCRI is a continuous variable, so it has better flexibility to quantify the

TABLE 4: The vehicle movement data and risk identification results for one car-following event.

Time interval (0.1 s)	FV speed (m/s)	FV speed – LV speed (m/s)	Headway distance (m)	Risk (TTC)	Risk (DRAC)	Risk (SDI)	Risk (RCRI)
...
158	16.9	9.4	38.7	0	0	1	0.08
159	16.7	9.7	37.9	0	0	1	0.10
160	16.4	10.2	36.5	0	0	1	0.16
161	16.1	10.4	35.7	0	0	1	0.18
162	15.9	10.5	34.9	0	0	1	0.21
163	15.6	10.7	33.3	0	1	1	0.27
164	15.2	10.7	32.5	0	1	1	0.30
165	14.9	10.8	31.7	1	1	1	0.31
166	14.6	11.2	30.0	1	1	1	0.43
...

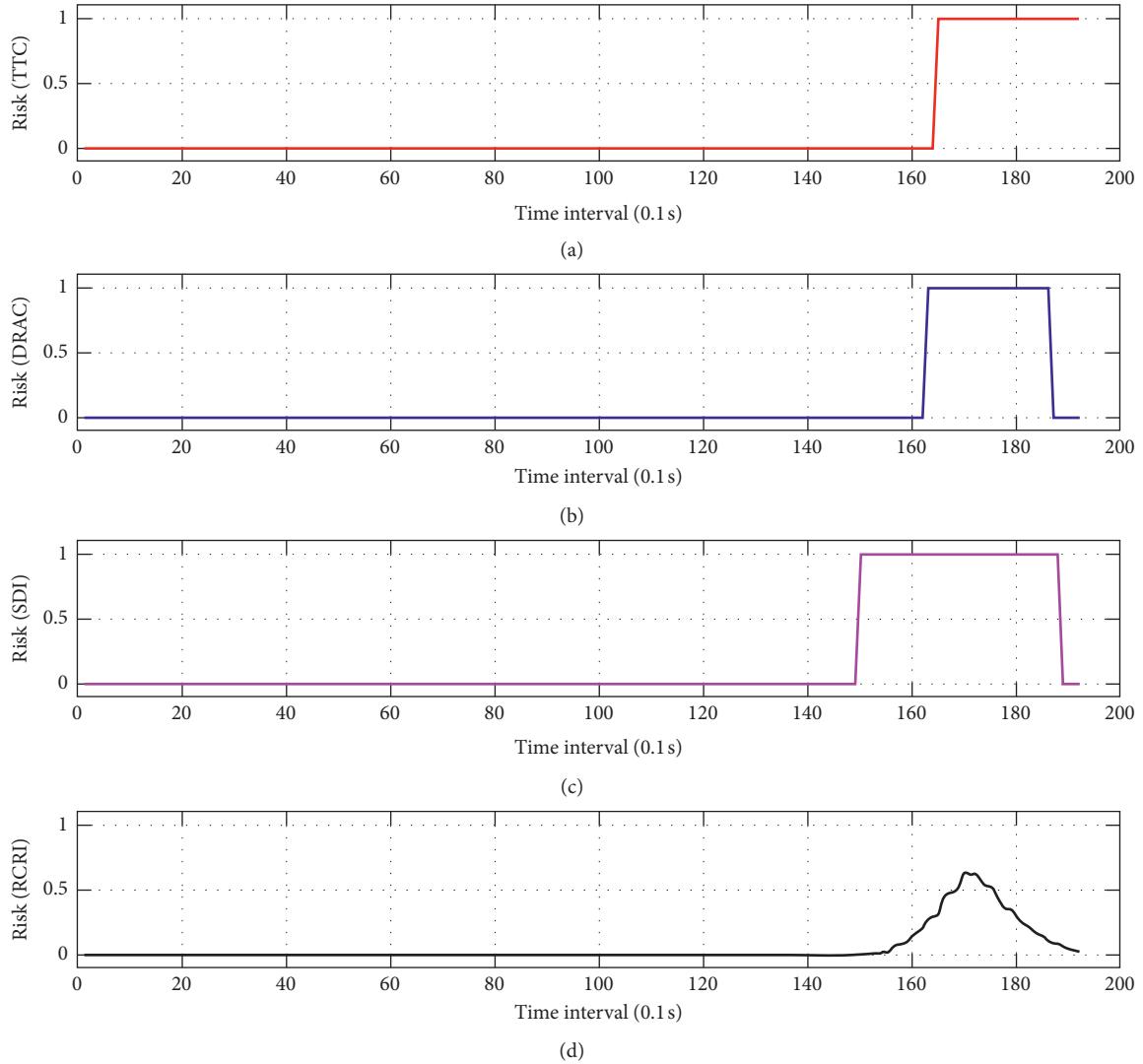


FIGURE 5: Rear-end crash risk identified by TTC, DRAC, SDI, and RCRI for one car-following event.

real-time change process of rear-end crash risk. On the contrary, the risk quantification results based on TTC, DRAC, and SDI are dummy variables. Therefore, the RCRI can more accurately represent the risk and has wider applicability.

5.2. Comparative Analysis of Rear-End Crash Risk under Different Influencing Factors. Before the significance test, the Kolmogorov-Smirnov (K-S) test was employed to verify the distribution of rear-end crash risks under different influencing factors. The results of the K-S test show that rear-end

TABLE 5: Significant main effects for rear-end crash risk.

Variables	Statistical results	Dummy variables	Rear-end crash risk	
			Mean	Std. Dev.
Day-of-week	$F = 58.1, p < 0.001$	Holiday	0.0028	0.0031
		Workday	0.0032	0.0033
Time-of-day	$F = 18.5, p < 0.001$	Off peak	0.0030	0.0033
		Morning peak	0.0034	0.0033
		Evening peak	0.0032	0.0032
Light condition	$F = 8.8, p = 0.003$	Daytime	0.0032	0.0032
		Nighttime	0.0030	0.0034
Weather condition	$F = 0.7, p = 0.408$	—	—	—
Traffic density	$F = 1655.0, p < 0.001$	High	0.0049	0.0027
		Median	0.0049	0.0033
		Low	0.0021	0.0029

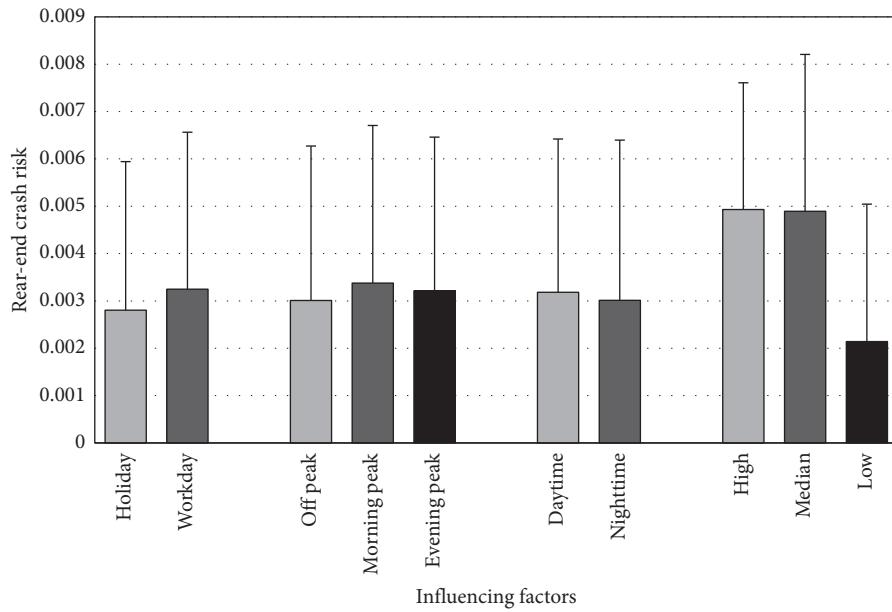


FIGURE 6: Rear-end crash risk under different influencing factors.

crash risk under different influencing factors meets the requirements of homogeneity of variance and normal distribution. Therefore, the analysis of variance (ANOVA) was then applied to test the significance of difference in rear-end crash risk under different influencing factors. Table 5 provides the summary statistics of the driving risk under these significant influencing factors. It can be seen from Table 5 that the rear-end crash risk under most influencing factors, such as day-of-week, time-of-day, light condition, and traffic density, is significantly different.

Figure 6 shows the comparison results of rear-end crash risk under different influencing factors. Specifically, for day-of-week, driving risk increased with the workday (by 14.3% from 0.0028 to 0.0032). In addition, the driving risks are 0.0030, 0.0034, and 0.0032 respectively in the three cases of off peak, morning peak, and evening peak. The significant differences in the risks of these temporal variables indicate that the driving risks are different for different travel

purposes. The decrease in driving risk was slight but significant by 0.0002 (6.3%) from daytime to nighttime. The driving risk decreased from 0.0049 in high traffic density to 0.0021 (by 57.1%) in low traffic density, indicating that traffic congestion leads to a decrease in driving safety.

5.3. Results of Mixed-Effects Linear Regression Model. Based on the car-following behavior variables and environmental factors variables obtained from SH-NDS, the influencing factors of rear-end crash risk are investigated. Table 6 presents the results of mixed-effects linear regression model. As shown in Table 6, the results of chi-squared goodness of fit indicate that the mixed-effects regression model fits well ($\chi^2(11) = 11198.24$, Prob $>\chi^2(11) = 0.00$). Except for the morning peak in temporal variables and weather conditions, all the variables listed in Table 6 are significant at 95% confidence level.

TABLE 6: Estimated parameters of the the mixed-effects linear regression model.

Explanatory variables	β	Std. Err.	Z	$P> Z $	95% CI			
					Lower	Upper		
<i>Fixed effects (behavioral variables)</i>								
Duration of car-following event	$-9.52e - 07$	$1.29e - 07$	-7.36	<0.001	$-1.20e - 06$	$-6.98e - 07$		
Mean time headway	-0.00200	0.00003	-65.32	<0.001	-0.00206	-0.00194		
Average speed difference	0.00029	$8.18e - 06$	35.07	<0.001	0.00027	0.00030		
Acceleration difference ratio	0.00099	0.00005	18.80	<0.001	0.00088	0.00109		
<i>Fixed effects (temporal variables)</i>								
Day-of-week	Holiday (baseline)	—	—	—	—	—		
	Workday	0.00022	0.00005	4.88	<0.001	0. 00013		
Time-of-day	Off peak (baseline)	—	—	—	—	—		
	Morning peak	0.00010	0.00005	1.91	0.056	$2.57e - 06$		
	Evening peak	-0.00012	0.00005	-2.42	0.016	-0.00023		
<i>Fixed effects (environmental variables)</i>								
Light condition	Daytime (baseline)	—	—	—	—	—		
	Nighttime	—	—	—	—	—		
Weather condition	Sunny (baseline)	—	—	—	—	—		
	Rainy	0.00013	0.00006	2.30	0.021	0.00002		
	High (baseline)	—	—	—	—	—		
Traffic density	Median	-0.00203	0.00007	-27.10	<0.001	-0.00217		
	Low	-0.00542	0.00007	-73.05	<0.001	-0.00557		
Fixed effects (intercept)	0.00915	0.00016	58.55	<0.001	0.00884	0.00946		
<i>Random effects (group name: driver)</i>								
		Estimate	Std. Err.	95% conf. interval				
Variance of intercept	5.53e - 07	$1.05e - 07$	$3.81e - 07$	$8.01e - 07$				
Variance of residual	$5.69e - 06$	$6.21e - 08$	$5.57e - 06$	$5.82e - 06$				
		$\chi^2(11) = 11198.24$, Prob $>\chi^2(11) = 0.0000$						
Log likelihood			77997.65					
LR test vs. linear model			$\chi^2 = 1216.65$	Prob $>\chi^2 = 0.00$				
Number of observations			16,905					
Number of groups			58					

As shown in Table 6, all selected behavioral variables affect the driving risk. The longer the duration of car-following event, the lower the driving risk. This can be understood as the risk of driving increases when the LV is changed frequently. Clearly, the larger the time headway, the lower the driving risk. This result is consistent with Duan et al. [9], who evaluated risk perception in car-following process. In addition, based on the SH-NDS data, Zhu et al. [43] concluded that the aggressive drivers have a shorter time gap than conservative drivers. Existing studies found evidence that speed dispersion is also an important factor in determining crash risk [44–46]. The larger speed difference between LV and FV is associated with a higher crash rate, which is generally consistent with our findings. In addition to the speed difference, this paper investigates the impact of acceleration difference on driving risk. The results show that the higher the acceleration difference ratio, the higher the risk, which indicates that the driving risk will increase when the FV uses more frequent acceleration and deceleration operations than the LV during car-following processes.

Qin et al. [47] suggested that due to the different travel purposes (to/from work) of drivers, the probability of a crash during working day and nonworking day is different, and the probability of crashes on working days is higher. This finding is consistent with the results of the regression model in this paper; that is, working days lead to higher driving risks. Furthermore, from the regression coefficients of this study, it

can be concluded that the crash risk is higher for the morning peak compared with other times of the day, and the evening peak is the lowest risk period of the day. The results are further confirmed that that there is a significant correlation between driving risk and driving purpose. The driving risk of drivers on the way to work is higher than the risk of leaving work.

As for the diverse environment, compared with sunny days, the risk of rear-end crash is higher for drivers on rainy days, which is consistent with Das et al. [48] and Jung et al. [49]. From the obtained results, we can draw the conclusion that traffic density has a greater impact on rear-end crash risk [50]. The variable of median-density and low-density show a negative coefficient ($\beta = -0.002$ for median-density and $\beta = -0.005$ for low-density), indicating that driving risk decreased in lower density traffic. The high-density traffic flow leads to an increase in the uncertainty of traffic flow and increases the driving risk. This result is consistent with Huang et al. [22], who investigated the driving risks under different conditions using naturalistic driving study and driver attitude questionnaire.

6. Conclusions

This study proposes a new SMoS to quantify driving risks in car-following situations and investigates the impact of different influencing factors (behavioral factors, temporal

factors, and environmental factors) on rear-end crash risk considering driver's heterogeneity. A total number of 16,905 car-following events were extracted from SH-NDS database. Risks of rear-end crash under different influencing factors were compared. In addition, a mixed-effects linear regression model was then applied to investigate the relationship between rear-end crash risk and various influencing factors. Several key conclusions can be drawn:

- (i) Different from TTC, DRAC, SDI, and other indicators, the surrogate measure RCRI was proposed based on crash mechanism and comprehensively considers the crash probability and consequences. This measure can be applied in any car-following situation, even when the speed of the LV is greater than the speed of the FV. The RCRI proposed in this study is a continuous variable, so it can be more flexible to quantify the risk of rear-end crash.
- (ii) Among different temporal variables, workday and morning peak hour had the highest mean value of driving risk. For different light conditions, the crash risk increased for daytime compared to nighttime. As for different traffic density, the driving risks corresponding to low-density traffic flows are significantly lower than those corresponding to high-density and medium-density traffic flows.
- (iii) The mixed-effects linear regression model performed well in quantitatively evaluating the impact of various influencing factors on rear-end crash risk. The developed models demonstrated that duration of car-following event, mean time headway, average speed difference, acceleration difference ratio, day-of-week, time-of-day, weather condition, and traffic density had significant effects on rear-end crash risk. Workday and morning peak negatively affected driver safety. As for environmental variables, rainy and high-density traffic decreased driver safety.

As the main contribution, this paper utilizes a new SMoS and naturalistic driving data to quantify rear-end crash risk and identify the impacts of different influencing factors on the crash risks. Research was conducted based on naturalistic driving data, which objectively reflects the real operation of drivers. The new SMoS can not only be used to investigate the driving safety of drivers under different driving environments but also be used for driving risk evaluation and real-time risk prediction. Results from the mixed-effects linear regression model can be used to improve driving safety by adopting appropriate countermeasures. For example, traffic safety management can be strengthened during working days and morning peak hours to ensure safe driving.

Still, limitations exist in this study. The proposed indicators mainly focus on the risk of rear-end crash and cannot comprehensively consider other types of crash risks. In addition, no crash data were obtained from SH-NDS database that can be used to verify the effectiveness of RCRI. For future work, further validation will be applied to evaluate the effectiveness of RCRI based on crash data.

Moreover, the RCRI will be used to predict the real-time change process of driving risk and explore the impact of risky driving.

Data Availability

The data used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This paper was jointly supported by the Chinese National Natural Science Foundation (71871161) and the Science and Technology Commission of Shanxi Province (19-JKCF-02).

References

- [1] WHO, *Global Status Report on Road Safety 2018*, World Health Organization, Geneva, Switzerland, 2018.
- [2] D. Tao, R. Zhang, and X. Qu, "The role of personality traits and driving experience in self-reported risky driving behaviors and accident risk among Chinese drivers," *Accident Analysis & Prevention*, vol. 99, pp. 228–235, 2017.
- [3] X. Zhang, X. Qu, D. Tao, and H. Xue, "The association between sensation seeking and driving outcomes: a systematic review and meta-analysis," *Accident Analysis & Prevention*, vol. 123, pp. 222–234, 2019.
- [4] S. Singh, *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, National Highway Traffic Safety Administration, Washington, DC, USA, 2015.
- [5] L. Dimitriou, K. Stylianou, and M. A. Abdel-Aty, "Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions," *Accident Analysis & Prevention*, vol. 118, pp. 221–235, 2018.
- [6] Y. Wu, M. Abdel-Aty, Q. Cai, J. Lee, and J. Park, "Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 11–25, 2018a.
- [7] National Highway Traffic Safety Administration, *Traffic Safety Facts 2015. Report No. Dot Hs 812 384*, National Highway Traffic Safety Administration, Washington, DC, USA, 2017.
- [8] K. L. M. Broughton, F. Switzer, and D. Scott, "Car following decisions under three visibility conditions and two speeds tested with a driving simulator," *Accident Analysis & Prevention*, vol. 39, no. 1, pp. 106–116, 2007.
- [9] J. Duan, Z. Li, and G. Salvendy, "Risk illusions in car following: is a smaller headway always perceived as more dangerous?" *Safety Science*, vol. 53, pp. 25–33, 2013.
- [10] X. Li, X. Yan, J. Wu, E. Radwan, and Y. Zhang, "A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender-A driving simulator based study," *Accident Analysis & Prevention*, vol. 97, pp. 1–18, 2016.
- [11] J. C. Hayward, "Near miss determination through use of a scale of danger," *Transportation Research Record*, vol. 384, pp. 24–34, 1972.

- [12] C. Oh, S. Park, and S. G. Ritchie, "A method for identifying rear-end collision risks using inductive loop detectors," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 295–301, 2006.
- [13] D. F. Cooper and N. Ferguson, "Traffic studies at t-junctions. 2. A conflict simulation record," *Traffic Engineering Control*, vol. 17, 1976.
- [14] G. Pnina, E. Johnathon, Z. Chunming et al., "Vehicle ownership and other predictors of teenagers risky driving behavior: evidence from a naturalistic driving study," *Accident Analysis & Prevention*, vol. 118, pp. 96–101, 2018.
- [15] B. Simons-Morton, N. Lerner, and J. Singer, "The observed effects of teenage passengers on the risky driving behavior of teenage drivers," *Accident Analysis & Prevention*, vol. 37, no. 6, pp. 973–982, 2005.
- [16] B. Yu, Y. Chen, S. Bao, and D. Xu, "Quantifying drivers' visual perception to analyze accident-prone locations on two-lane mountain highways," *Accident Analysis & Prevention*, vol. 119, pp. 122–130, 2018.
- [17] Y. Wu, M. Abdel-Aty, J. Park, and R. M. Selby, "Effects of real-time warning systems on driving under fog conditions using an empirically supported speed choice modeling framework," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 97–110, 2018.
- [18] Y. Kuang, X. Qu, and S. Wang, "A tree-structured crash surrogate measure for freeways," *Accident Analysis & Prevention*, vol. 77, pp. 137–148, 2015.
- [19] K. Xie, D. Yang, K. Ozbay, and H. Yang, "Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure," *Accident Analysis & Prevention*, vol. 125, pp. 311–319, 2019.
- [20] X. Shi, Y. D. Wong, M. Z. F. Li, and C. Chai, "Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory," *Accident Analysis & Prevention*, vol. 117, pp. 346–356, 2018.
- [21] T. Fu, L. Miranda-Moreno, and N. Saunier, "A novel framework to evaluate pedestrian safety at non-signalized locations," *Accident Analysis & Prevention*, vol. 111, pp. 23–33, 2018.
- [22] H. Huang, Y. Li, X. Zheng, J. Wang, Q. Xu, and S. Zheng, "Objective and subjective analysis to quantify influence factors of driving risk," in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 4310–4316, IEEE, Auckland, New Zealand, October 2019.
- [23] A. C. Hayley, B. d. Ridder, C. Stough, T. C. Ford, and L. A. Downey, "Emotional intelligence and risky driving behaviour in adults," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 49, pp. 124–131, 2017.
- [24] E. L. Harbeck, A. I. Glendon, and T. J. Hine, "Young driver perceived risk and risky driving: a theoretical approach to the "fatal five,'" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 58, pp. 392–404, 2018.
- [25] P. Ulleberg and T. Rundmo, "Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers," *Safety Science*, vol. 41, no. 5, pp. 427–443, 2003.
- [26] E. Teye-Kwadjo, "Risky driving behaviour in urban Ghana: the contributions of fatalistic beliefs, risk perception, and risk-taking attitude," *International Journal of Health Promotion and Education*, vol. 57, no. 5, pp. 256–273, 2019.
- [27] Q. Shangguan, T. Fu, and S. Liu, "Investigating rear-end collision avoidance behavior under varied foggy weather conditions: a study using advanced driving simulator and survival analysis," *Accident Analysis & Prevention*, vol. 139, Article ID 105499, 2020.
- [28] L. Precht, A. Keinath, and J. F. Krems, "Identifying the main factors contributing to driving errors and traffic violations—results from naturalistic driving data," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 49, pp. 49–92, 2017.
- [29] Q. Chen, R. Gu, H. Huang, J. Lee, X. Zhai, and Y. Li, "Using vehicular trajectory data to explore risky factors and unobserved heterogeneity during lane-changing," *Accident Analysis & Prevention*, vol. 151, Article ID 105871, 2021.
- [30] X. Wang, M. Yang, and D. Hurwitz, "Analysis of cut-in behavior based on naturalistic driving data," *Accident Analysis & Prevention*, vol. 124, pp. 127–137, 2019.
- [31] M. Zhu, X. Wang, and J. Hu, "Impact on car following behavior of a forward collision warning system with headway monitoring," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 226–244, 2020.
- [32] J. Wang, T. Luo, and T. Fu, "Crash prediction based on traffic platoon characteristics using floating car trajectory data and the machine learning approach," *Accident Analysis & Prevention*, vol. 133, Article ID 105320, 2019.
- [33] M. Zhu, X. Wang, A. Tarko, and S. E. Fang, "Modeling car-following behavior on urban expressways in shanghai: a naturalistic driving study," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 425–445, 2018.
- [34] C. Johnsson, A. Laureshyn, and T. De Ceunynck, "In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators," *Transport Reviews*, vol. 38, no. 6, pp. 765–785, 2018.
- [35] J. Carbaugh, D. N. Godbole, and R. Sengupta, "Safety and capacity analysis of automated and manual highway systems," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 1-2, pp. 69–99, 1998.
- [36] J. Wang, Y. Kong, T. Fu, and J. Stipancic, "The impact of vehicle moving violations and freeway traffic flow on crash risk: an application of plugin development for micro-simulation," *PLoS One*, vol. 12, 2017.
- [37] D. Wang, *Traffic Flow Theory*, China Communications Press, Beijing, China, 2002.
- [38] F. Cunto and F. F. Saccomanno, "Calibration and validation of simulated vehicle safety performance at signalized intersections," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1171–1179, 2008.
- [39] M. Yang, X. Wang, and M. Quddus, "Examining lane change gap acceptance, duration and impact using naturalistic driving data," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 317–331, 2019.
- [40] H. Gan and Y. Bai, "The effect of travel time variability on route choice decision: a generalized linear mixed model based analysis," *Transportation*, vol. 41, no. 2, pp. 339–350, 2014.
- [41] C. Chai, X. Zeng, X. Wu, and X. Wang, "Safety evaluation of responsibility-sensitive safety (RSS) on autonomous car-following maneuvers based on surrogate safety measurements," in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 175–180, IEEE, Auckland, New Zealand, October 2019.
- [42] G. Guido, F. Saccomanno, A. Vitale, V. Astarita, and D. Festa, "Comparing safety performance measures obtained from video capture data," *Journal of Transportation Engineering*, vol. 137, no. 7, pp. 481–491, 2011.
- [43] M. Zhu, X. Wang, and Y. Wang, "Human-like autonomous car-following model with deep reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 348–368, 2018.

- [44] L. Aarts and I. Van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [45] O. M. J. Carsten and F. N. Tate, "Intelligent speed adaptation: accident savings and cost-benefit analysis," *Accident Analysis & Prevention*, vol. 37, no. 3, pp. 407–416, 2005.
- [46] T. Stübig, M. Petri, C. Zeckey et al., "Alcohol intoxication in road traffic accidents leads to higher impact speed difference, higher ISS and MAIS, and higher preclinical mortality," *Alcohol*, vol. 46, no. 7, pp. 681–686, 2012.
- [47] X. Qin, J. N. Ivan, N. Ravishanker, J. Liu, and D. Tepas, "Bayesian estimation of hourly exposure functions by crash type and time of day," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1071–1080, 2006.
- [48] S. Das, A. Dutta, and S. R. Geedipally, "Applying Bayesian data mining to measure the effect of vehicular defects on crash severity," *Journal of Transportation Safety & Security*, pp. 1–17, 2019.
- [49] S. Jung, X. Qin, and D. A. Noyce, "Modeling highway safety and simulation in rainy weather," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2237, no. 1, pp. 134–143, 2011.
- [50] Y. Guo, T. Sayed, and M. Essa, "Real-time conflict-based Bayesian Tobit models for safety evaluation of signalized intersections," *Accident Analysis & Prevention*, vol. 144, Article ID 105660, 2020.