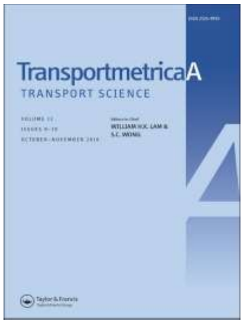

Taylor & Francis
Taylor & Francis Group

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

A matched case-control method to model car-following safety

Qianwen Li, Handong Yao & Xiaopeng Li

To cite this article: Qianwen Li, Handong Yao & Xiaopeng Li (2022): A matched case-control method to model car-following safety, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2022.2055198

To link to this article: https://doi.org/10.1080/23249935.2022.2055198

Published online: 25 Mar 2022.

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2022.2055198

Taylor & Francis
Taylor & Francis Group

A matched case-control method to model car-following safety

Qianwen Li, Handong Yao and Xiaopeng Li

Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL, USA

ABSTRACT
Car-following safety is related to both observed driving characteristics (e.g. car-following behaviour) and unobserved driver heterogeneity (e.g. drivers' psychological features). Two major issues remain in the existing literature, i.e. limiting to longitudinal characteristics and not addressing the confounding effects of unobserved driver heterogeneity. This study takes a matched case–control approach to model car-following safety with both longitudinal and lateral driving characteristics. Unobserved driver heterogeneity is controlled by matching preceding and following vehicle IDs. Results show that unstable lateral movements of preceding vehicles and following vehicles contribute to higher crash risks. Comparison results on two datasets with different congestion levels reveal that it is safer in more congested traffic when the following vehicle maintains more stable longitudinal and lateral behaviours, and greater speed difference, headway, and spacing regarding its preceding vehicle. This study provides insights in enhancing roadway safety management and benefiting the automated vehicle development by warnings on associated risks.

ARTICLE HISTORY
Received 13 October 2021
Accepted 13 March 2022

KEYWORDS
Car-following safety; driver heterogeneity; longitudinal and lateral movements; trajectory data; automated vehicle safety warning

Introduction

The World Health Organization (WHO) reported on 21st June 2021 that approximately 1.3 million people die each year because of road traffic crashes. Especially for children and young adults aged 5–29 years, road traffic crashes are the leading cause of death (WHO 2021). To reduce road traffic crashes, great efforts have been made to model roadway safety and investigate contributing risk factors. This not only enhances traffic safety management by guiding stakeholders and policymakers to devise proper safety guidelines and regulations but also facilitates automated vehicle development by warning on risky situations.

Automated vehicle technology has witnessed substantial development in the past decades. It was predicted that there would be 20.8 million automated vehicles in operations in the US by 2030 (Statista 2017). Despite the promising growth, most people still hold concerns about the safety performance of automated vehicles (Koopman and Wagner 2017), especially given the increasing number of automated vehicle crashes (Blanco et al. 2016; Song, Chitturi, and Noyce 2021). To facilitate automated vehicle development,

CONTACT Handong Yao handongyao@usf.edu

safety concerns must be addressed first. Enabled by various onboard sensors (e.g. Lidar, Radar, and cameras), automated vehicles percept the movements of the surrounding traffic (e.g. vehicle speeds and positions). With these perceptions, automated vehicles can evaluate the current safety performance and respond correspondingly to avoid any potential crashes if any, i.e. adopt a deceleration. With the hardware ready, the next necessary component to develop a real-time automated vehicle safety warning system is an effective safety evaluation model based on the sensed surrounding vehicle motions.

In the past, vehicle motion data have been widely used to assess crash risks when historical crash data is not available. Various safety surrogate measures (SSMs) were developed in the past, e.g. time to collision (TTC), deceleration rate to avoid crash (DRAC), and stop distance index (SDI), to measure traffic safety with vehicle trajectory data (Peng, Lyu, and Wu 2020; Shi et al. 2018).

With these efficient SSMs, researchers have studied the crash risks of two main driving behaviours. The first one is lane-changing behaviour, which yields a high crash risk due to the complicated vehicle interactions (Chen et al. 2021a, 2021b; Gu et al. 2019; Park et al. 2018). Another one is car-following behaviour, which is the most common action in a traffic stream and might cause rear-end crashes due to traffic flow instability (Behbahani, Nadimi, and Naseralavi 2015; Li et al. 2020; Przybyla et al. 2015; Xue et al. 2019; Zhao and Lee 2018). This paper focuses on the crash risks of car-following behaviour. While the existing car-following studies laid a solid foundation for understanding risky traffic scenarios, two main issues are still unaddressed. First, most of them focused on longitudinal characteristics, e.g. longitudinal speed and position. Lateral characteristics have not been paid much attention, despite their non-neglectable impacts. Yu, Han, and Zhang (2021) only considered the maximum lateral acceleration of the preceding vehicle in high-risk events prediction. Since vehicles with difficulties in lane-keeping, deviating from the lane centre frequently, are less stable and are expected to have a greater possibility to experience unsafe scenarios (Zhang, Dong, and Du 2008), more lateral characteristics should be investigated for a comprehensive understanding of car-following safety.

Most importantly, the existing literature did not properly handle the safety impacts of the unobserved driver heterogeneity (e.g. mental characteristics) when investigating observed driving characteristics (e.g. car-following). As confirmed by the existing literature, driver heterogeneity significantly pertains to rear-end crash risks because different drivers respond to traffic scenarios differently (James et al. 2019; Xie et al. 2019; Yu, Han, and Zhang 2021; Zhang, Wang, and Lu 2019). In this case, unobserved driver heterogeneity is a confounding variable of other driving characteristics when modelling car-following safety. It is necessary to control it to derive unbiased estimation results of the safety effects of other risk factors that can be easily observed and thus readily implemented in safety management.

The matched case–control is a promising solution to address the confounding effects of unobserved driver heterogeneity. It has been widely used in epidemiological research (Breslow 1996; Breslow et al. 1978; Ingram et al. 1997; Wey et al. 1989). It has also witnessed substantial development in roadway safety modelling crashes (Gjerde et al. 2011; Gross 2013; Híjar et al. 2000; Khoda Bakhshi and Ahmed 2021). It categorises entities into two groups, i.e. cases (unsafe entities) and controls (safe entities). Each case is randomly matched to several controls by the confounding variable to form a stratum. In this study, each entity is a trajectory segment, and unobserved driver heterogeneity is controlled by matching preceding vehicle ID and following vehicle ID. We assume that unobserved driver-specific characteristics are the same in a stratum based on the premise that these characteristics do not change much within a short travel period. After matching, the confounding effects of unobserved driver heterogeneity are eliminated and thus unbiased safety effects of observed driving characteristics can be investigated. A conditional logistic regression model is estimated on the matched data to quantify the safety effects of observed driving characteristics on car-following safety.

The contributions of this study are summarised as follows.

(1) For the first time, this study uses the matched case–control method to control the confounding effects of unobserved driver heterogeneity and quantify the car-following safety effects of observed driving characteristics.
(2) This study assesses the car-following safety impacts of both lateral and longitudinal driving characteristics.
(3) This study compares the model results of two trajectory datasets with different congestion levels (one severely congested and one slightly congested).

The study structure is depicted in Figure 1. Vehicle car-following trajectories are extracted and segmented from original trajectory datasets. The safety performance of each car-following segment is assessed by SSM. A segment is labelled as a case if it experiences crash risks and labelled as a control if it does not experience crash risks. After labelling, each case is randomly matched to several controls by preceding vehicle ID and following vehicle ID to eliminate the confounding effects of unobserved driver heterogeneity. A group of one case and several matched controls is named a stratum. The conditional logistic regression model is estimated on the matched strata to quantify the safety effects of

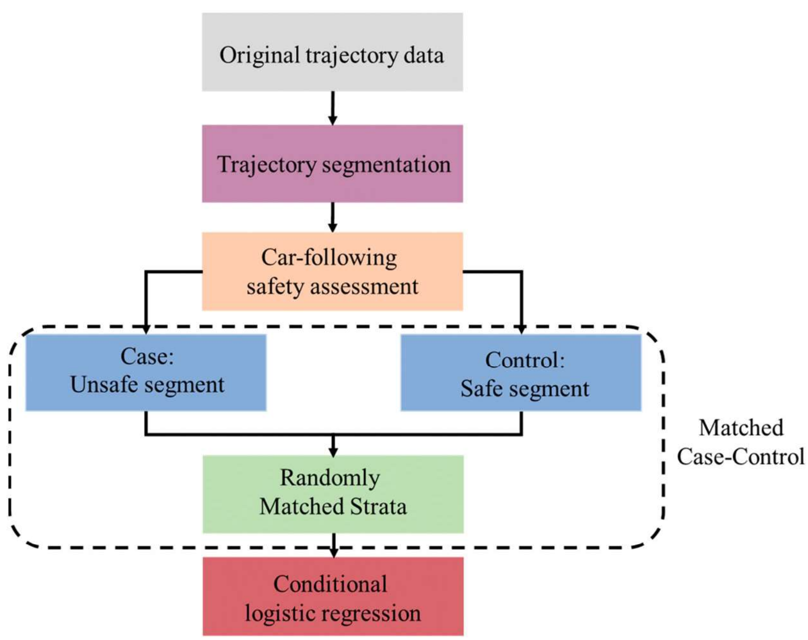

Figure 1. Study structure.

observed driving characteristics after controlling the confounding effects of unobserved driver heterogeneity.

The rest of this paper is organised as follows. Section 2 introduces the car-following safety assessment, the matched case–control study, and the conditional logistic regression. Section 3 describes empirical settings including two trajectory datasets (one severely congested and one slightly congested) and the data processing. Section 4 discusses the model estimation results. Section 5 concludes this study with ending remarks.

Methodology

This section introduces the car-following safety index, describes the matched case–control study, and reviews the conditional logistic regression.

Car-following safety assessment

The stopping distance index (SDI) is adopted to assess car-following safety in this study. It compares the safe stopping distances of two vehicles for risk assessments (Chen et al. 2021a; Oh and Kim 2010; Park et al. 2018).

First, the safe stopping distances at time *t* are calculated for the preceding vehicle ($\mathrm{SSD_P}$) and the following vehicle ($\mathrm{SSD_F}$).

$$\mathrm{SSD_P}(t) = \frac{v_P^2(t)}{2a_P}. \tag{1}$$

$$\mathrm{SSD_F}(t) = v_F(t) \times \tau + \frac{v_F^2(t)}{2a_F}. \tag{2}$$

where $v_\mathrm{P}(t)$ and $v_\mathrm{F}(t)$ are the speeds of the preceding vehicle and the following vehicle at time t; τ is the perception reaction time of the following vehicle and set as 1.5 s (Chen et al. 2021a; Oh and Kim 2010; Park et al. 2018); and a_P and a_F are the maximum deceleration rates of the preceding vehicle and the following vehicle (3.4 m/s^2 for passenger cars, 2.4 m/s^2 for trucks, and 4.5 m/s^2 for motorcycles) (AASHTO 2001; Chen et al. 2021b; Huertas-Leyva et al. 2019; Kweon 2011; Wu et al. 2018a). The SDI of the following vehicle is calculated.

$$\mathrm{SDI}(t) = \begin{cases} 0, & \text{if } \mathrm{SSD_P}(t) + sp(t) - \mathrm{SSD_F}(t) \geq 0, \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

where $sp(t)$ is the spacing between the preceding vehicle's rear bumper and the following vehicle's front bumper at time t.

Other popular SSMs can also be adopted to evaluate car-following safety, e.g. TTC, DRAC, and post encroachment time (PET). The methodology proposed below remains valid regardless of the specific SSM. It should be highlighted that the SSM selection in terms of devising a vehicle safety warning system should be carefully decided after extensive testing on various SSMs. The warning system should be neither too aggressive (missing certain risky situations) nor too conservative (warning 'fake' risky situations). The trade-off between the system effectiveness (not missing any risks) and accuracy (reporting only 'true' risks) should be balanced per application needs when selecting SSMs and setting thresholds for continuous SSMs (e.g. TTC) in differentiating risky from safe.

For example, if the warning system is designed for human-driven vehicles or automated vehicles of lower automation levels (SAE Level 3 and below), a less conservative SSM should be considered because too frequent warnings may irritate the driver who is (partially) operating the vehicle. As a result, he/she might just turn off the system, resulting in safety concerns (Jamson, Lai, and Carsten 2008). However, if the warning system is intended for automated vehicles of higher automation levels (SAE Level 4 and up), it can be more conservative for a higher level of safety. Since the vehicle is operating itself instead of a driver, the warning messages will be sent to the vehicle in silence instead of irritating the rider who is enjoying the onboard time with activities other than driving. In this paper, except for the SDI-based results reported in Section 4, the TTC-based results are also presented in Appendix.

Matched case–control method

The matched case–control method has been well established in epidemiology to investigate risk factors to disease since decades ago (Breslow 1996; Breslow et al. 1978; Ingram et al. 1997; Wey et al. 1989). It has also been used in transportation to study the contributing factors to traffic crashes (Gjerde et al. 2011; Gross 2013; Híjar et al. 2000). The following three steps are required when conducting a matched case–control study (Schlesselman 1982).

(1) Defining cases and controls. Cases are unsafe car-following segments that experience crash risks and controls are safe car-following segments that do not experience crash risks.
(2) Randomly matching cases and controls. Each case is randomly matched to several controls to form a stratum based on confounding variables (i.e. unobserved driver characteristics) that are related to both risk factors of interest (i.e. observed driving characteristics) and the outcome (i.e. unsafe car-following). After matching, the impacts of confounding variables on the outcome are eliminated and thus unbiased associations between risk factors of interest and the outcome can be estimated. As the matching ratio (i.e. control to case) increases, the analysis power of the matched case–control study increases. Yet, the number of resulting strata decreases. Thus, the matching ratio shall be carefully chosen by balancing the two aspects.
(3) Conducting analysis. Conditional logistic regression was developed to fit matched data in case–control studies. It is an extension of logistic regression (Breslow et al. 1978).

Conditional logistic regression

The probability in the jth observation of the ith stratum being a case (i.e. unsafe car-following segment) is calculated as follows.

$$\Pr(y_{ij} = 1) = \frac{\exp(\alpha_i + \boldsymbol{\beta X_{ij}})}{1 + \exp(\alpha_i + \boldsymbol{\beta X_{ij}})} \tag{4}$$

where $\boldsymbol{X_{ij}}$ is the vector of explanatory variables (i.e. risk factors) associated with y_{ij}; $\boldsymbol{\beta}$ is the vector of coefficients corresponding to $\boldsymbol{X_{ij}}$; αi is the stratum-specific interpretation term reflecting the different combination effects of confounding variables for different strata.

The conditional likelihood of stratum i is calculated as follows.

$$L(Y_i|\boldsymbol{\beta}) = P\left(y_{i1} = 1, y_{ij} = 0 \quad \text{for} \quad j > 1 \mid \boldsymbol{X_{ij}}, \sum_{j=1}^{J} y_{ij} = 1, \boldsymbol{\beta}\right) = \frac{\exp(\boldsymbol{\beta X_{i1}})}{\sum_{j\in J}\exp(\boldsymbol{\beta X_{ij}})} \quad (5)$$

where y_{i1} is the case observation of the ith stratum; y_{ij} when $j > 1$ are the matched controls of the ith stratum; $\boldsymbol{X_{ij}}$ is the vector of explanatory variables associated with y_{ij}. When the strata are independent of each other, the conditional log-likelihood function $(Y|\boldsymbol{\beta})$ over I strata is calculated in Equation (6) (Schlesselman 1982).

$$LL(Y|\boldsymbol{\beta}) = -\sum_{i=1}^{I} \ln\left\{1 + \sum_{j\epsilon J}\exp[\boldsymbol{\beta}(\boldsymbol{X_{ij}} - \boldsymbol{X_{i1}})]\right\} \quad (6)$$

The Maximum Likelihood Estimation is used to maximise $LL(Y|\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$.

The fitted conditional logistic regression model yields the odds ratio. It is measured to assess the effects of one unit change in unmatched risk factors on unsafe car-following. The odds ratio (OR) for a continuous factor x_m is computed as in Equation (7).

$$\text{OR}(x_m) = \frac{\Pr(y_{i1} = 1|x_m = x', \boldsymbol{Z})/[1 - \Pr(y_{i1} = 1|x_m = x', \boldsymbol{Z})]}{\Pr(y_{i1} = 1|x_m = x' + \Delta, \boldsymbol{Z})/[1 - \Pr(y_{i1} = 1|x_m = x' + \Delta, \boldsymbol{Z})]} = \exp(\beta_m) \quad (7)$$

where Z is the vector of risk factors other than x_m; x' is the old value of x_m; Δ is the changes in x_m; and β_m is the estimated parameter for x_m.

Empirical settings

This section first introduces two vehicle trajectory datasets and then describes the data processing.

Trajectory dataset

Two vehicle trajectory datasets are adopted in this study. The first dataset is the widely used next generation simulation (NGSIM) data collected by the Federal Highway Administration on a segment of U.S. Highway 101 in Los Angeles, California between 7:50 am to 8:05 am on June 15, 2005 (FHWA 2008). Video data were collected with eight cameras. The US101 segment is 2100 feet long with five main lanes, an auxiliary lane, an on-ramp, and an off-ramp. The data frequency is 10 Hz.

The second one is the high-granularity highway simulation (HIGHSIM) vehicle trajectory dataset collected by the Federal Highway Administration. Aerial videos were recorded by three 8 K cameras on three helicopters from 4:15 pm to 6:15 pm on May 14, 2019 (Tuesday) on a segment of the Interstate 75 (I-75) in Florida, USA, shown in Figure 2. The I-75 segment is 8000 ft long with three main lanes and an off-ramp. The Video-Based Intelligent Road Traffic Universal Analysis Tool (VIRTUAL) developed by the University of South Florida was used to extract trajectories from aerial videos (Shi et al. 2021). The format of the HIGHSIM I-75 dataset is similar to that of the NGSIM US101 dataset, including vehicle ID, time, longitudinal position, lateral position, width, length, speed, acceleration, lane number, space headway, and vehicle class (i.e. truck, passenger car, or motorcycle). The data frequency is 30 Hz. To

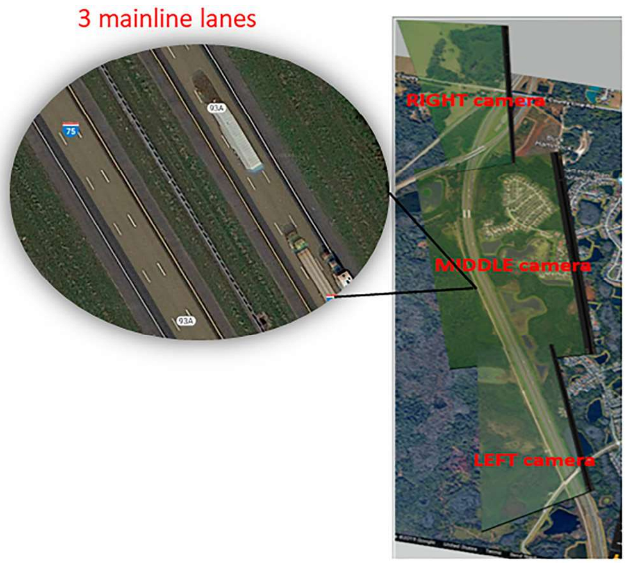

Figure 2. Data collection segment on the I-75 freeway.

be consistent with the NGSIM US101 dataset, the HIGHSIM I-75 dataset is resampled to the 10 Hz data frequency. For more information, please refer to Shi et al. (2021).

To illustrate the traffic operations of the two datasets, the flow-density diagrams are constructed. When car-following trajectory data is available, Edie's equations (Edie 1963) can be used to calculate macroscopic traffic variables (e.g. flow and density) with a subset of trajectories in an arbitrary time–space region. For more information about the detailed calculation, please refer to (Knoop et al. 2012; Laval 2011).

The flow-density diagrams of the two datasets are plotted in Figure 3. It is clearly shown that the US101 traffic is more congested. All data points are located in the congested region with the density ranging from 20 vehicles per kilometre per lane to 90 vehicles per kilometre per lane, shown in Figure 3(a). The I-75 traffic is less congested and the density is less than

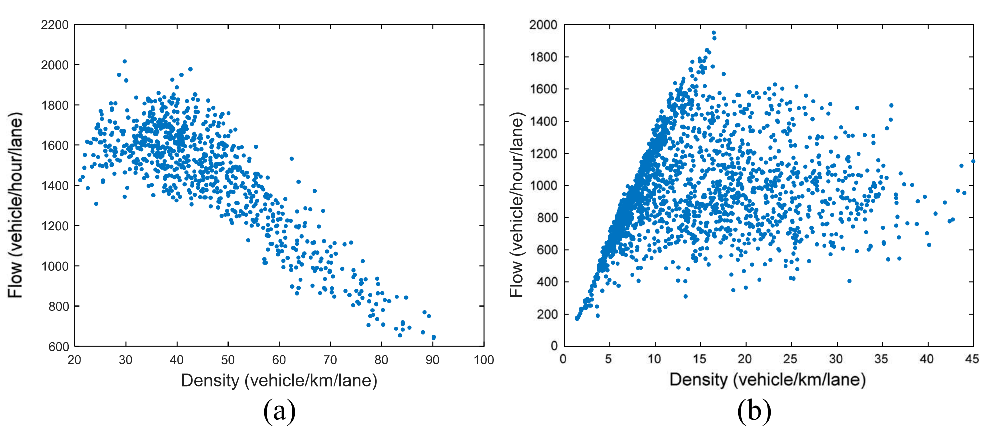

Figure 3. Flow-density diagrams; (a) US101 dataset; (b) I-75 dataset.

Table 1. Basic statistics of the original datasets.

	Speed (m/s)			Spacing (m)			Observation period of car-following pairs (s)	
Dataset	Min	Max	Mean	Min	Max	Mean	Max	Mean
NGSIM (US101)	0	29.39	9.36	0	239.85	20.78	123.40	35.85
HIGHSIM (I-75)	0	46.26	18.26	0.004	734.06	57.78	368.93	36.30

45 vehicles per kilometre per lane, shown in Figure 3 (b). A significant number of data points are located in the free flow region. Table 1 presents some basic statistics of the two original datasets. The less congested I-75 traffic from the HIGHSIM I-75 dataset has higher speed and longer spacing than the US101 traffic from the NGSIM US101 dataset. Both two datasets have similar observation periods of car-following pairs, i.e. around 36 s. The car-following pair refers to the combination of a preceding vehicle and the following vehicle. As long the two consecutive vehicles remain the same, the car-following pair exists; otherwise, the car-following pair disappears.

Data processing

Since vehicles in ramps usually have different car-following behaviour (e.g. slower because of the road geometry design and mandatory lane changes) and much shorter observation periods compared with those in mainline, we only consider vehicle trajectory data in the mainline to investigate the car-following behaviour.

Risk factors are extracted as follows. First, longitudinal speed, longitudinal acceleration, lateral position, spacing, and headway are extracted from datasets directly. The longitudinal speed difference is calculated as the longitudinal speed of the preceding vehicle minus that of the following vehicle. The following vehicle average deviation from the lane centre is calculated as the absolute value of the difference between the vehicle lateral position and the lane centre. The following vehicle and preceding vehicle lateral speeds are computed as the absolute values of the first-order differential of their lateral positions. Greater deviation or greater lateral speed means more laterally unstable. Descriptive statistics of risk factors are provided in Table 2.

Next, car-following safety observations are decided. Vehicle trajectories of each car-following pair are segmented by a length of 5 s. Each segment represents an observation. The length of 5 is chosen for the following reasons. This segment can neither be too long nor too short. If it is too long, given the short overall car-following period (see Table 3, about 35 s on average), the resulting sample size would greatly decrease. For example, when the segmentation length is extended from 5 s to 10 s, the resulting sample size of the NGSIM US101 is reduced to only 25%. Further, 5-s of trajectory data has captured a relatively stable car-following behaviour like the 10-s data. Testing results reveal that the OR values when 5-s data are used are similar to those when 10-s data are used.

On the other hand, if the segment is too short, it would not be able to characterise a stable car-following behaviour (Lu, Varaiya, and Horowitz 2009). The limited car-following trajectory data is not enough to capture the vehicle movement variation. The risk factor values of cases are similar to the values of controls, resulting in insignificant factors. For example, when the segment is set as only 1 s, the z statistic of diff_vstd in the I-75 model is as low as 0.73 and the corresponding P value is as high as 0.464; the z statistic of FV_vstd

Table 2. Descriptive statistics of explanatory variables.

	US101				I-75			
	Case ($n = 4564$)		Control ($n = 9128$)		Case ($n = 2157$)		Control ($n = 4314$)	
Variable description	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Interaction risk factors								
diff_vmn: Longitudinal average speed difference between the preceding vehicle and following vehicle (m/s)	−0.45	1.03	0.66	1.08	−1.29	2.13	0.31	1.55
diff_vstd: Longitudinal speed difference standard deviation (m/s)	0.88	0.47	0.73	0.37	0.48	0.52	0.29	0.32
sp_mn: Average spacing (m)	14.71	7.21	18.39	10.34	36.54	23.37	38.10	31.00
hw_mn: Average headway (s)	2.08	2.01	3.33	3.81	1.78	0.81	3.04	2.33
Individual vehicle risk factors-longitudinal factors								
FV_vmn: Following vehicle longitudinal average speed (m/s)	8.57	3.47	7.39	4.21	21.70	9.61	15.53	11.23
PV_amax: Preceding vehicle longitudinal maximum acceleration (m/s^2)	1.64	0.93	1.94	0.88	2.04	0.99	2.51	0.74
FV_vstd: Following vehicle longitudinal speed standard deviation (m/s)	0.82	0.50	0.73	0.47	0.81	0.77	0.45	0.40
Individual vehicle risk factors-lateral factors								
FV_xmn: Following vehicle average deviation from the lane centre (m)	0.37	0.22	0.33	0.22	0.31	0.22	0.35	0.24
FV_vx_mn: Following vehicle average lateral speed (0.1 m/s)	1.17	0.82	0.87	0.72	0.49	0.64	0.38	0.37
PV_vx_mn: Preceding vehicle average lateral speed (0.1 m/s)	1.16	0.83	0.93	0.75	0.42	0.49	0.39	0.43

in the US101 model is as low as 0.59 and the corresponding P value is as high as 0.558. This loses the opportunity to identify some intuitive risk factors, e.g. speed oscillation.

We also want to note that the segment length is dataset sensitive. Values are valid as long as they characterise a stable car-following behaviour while guaranteeing the sample size.

Segments satisfying the following criteria (1)–(2) are excluded. After selecting suitable car-following segments, a linear interpolation is conducted to complete trajectories with missing data points. And the Savitzky–Golay filter is used on speed and acceleration to reduce noises.

(1) Maximum deceleration goes beyond the SSD calculation thresholds, i.e. 3.4 m/s^2 for passenger cars, 2.4 m/s^2 for trucks, and 4.5 m/s^2 for motorcycles (Chen et al. 2021b).
(2) The average spacing is greater than 300 m. In this case, the following vehicle is not much affected by the preceding vehicle and thus the concept of car-following does not exist.

The safety of each car-following segment is assessed as follows. If the SDI value, calculated in Equation (3), is always equal to 0, this segment is defined as safe (i.e. control).

Table 3. Fitted conditional logistic model results.

	US101						I-75					
Variable	Coef.	***z***	*p*-value	OR	95% CI of OR	SE of OR	Coef.	***z***	*p*-value	OR	95% CI of OR	SE of OR
Interaction risk factors												
diff_vmn (m/s)	−1.565	−33.99	0.000	0.209	[0.191, 0.229]	0.010	−0.730	−19.00	0.000	0.482	[0.447, 0.519]	0.019
diff_vstd (m/s)	1.703	17.91	0.000	5.493	[4.559, 6.619]	0.522	0.328	3.95	0.000	1.388	[1.301, 2.189]	0.224
sp_mn (m)	−0.165	−26.76	0.000	0.848	[0.838, 0.859]	0.005						
hw_mn (s)	−0.660	−20.78	0.000	0.517	[0.486, 0.550]	0.016						
Individual vehicle risk factors-longitudinal factors												
FV_vmn (m/s)							0.146	18.62	0.000	1.157	[1.139, 1.175]	0.009
PV_amax (m/s^2)	−0.088	−2.23	0.026	0.916	[0.848, 0.990]	0.036	−0.182	−3.80	0.000	0.834	[0.759, 0.916]	0.040
FV_vstd (m/s)	0.655	7.94	0.000	1.925	[1.638, 2.264]	0.158	0.817	8.47	0.000	2.264	[1.874, 2.736]	0.219
Individual vehicle risk factors-lateral factors												
FV_xmn (m)	0.998	6.90	0.000	2.712	[2.043, 3.601]	0.392	0.487	2.25	0.024	1.627	[1.065, 2.484]	0.351
FV_vx_mn (0.1 m/s)	0.449	9.49	0.000	1.566	[1.428, 1.718]	0.074	0.204	2.02	0.043	1.226	[1.006, 1.493]	0.123
PV_vx_mn (0.1 m/s)	0.568	11.97	0.000	1.764	[1.607, 1.936]	0.084						
Model Statistics												
Number of observations	13,692						6,471					
Log-likelihood	−1926.26						−1230.22					
Pseudo R^2	0.616						0.481					

Otherwise, it is defined as unsafe (i.e. case), indicating there are risks that the following vehicle collides with the preceding vehicle. According to Equation (3), the SDI value is calculated at each time step (i.e. 0.1 s). Considering a 5-s car-following segment, there are 50 SDI values. A safe car-following segment is identified when all 50 SDI values are 0. Risks exist when at least one SDI value equals 1, and thus this car-following segment is unsafe.

To control unobserved driver-specific characteristics, each case is randomly matched to two controls by preceding vehicle ID and following vehicle ID to form a stratum. The matching ratio of two is chosen after balancing the sample size and the analysis power. With two controls matched to each case, the analysis power reaches about 96% (Woodward 2013). By conducting such matching, the confounding impacts of unobserved driver heterogeneity can be eliminated because the driver is the same in each stratum. The premise is that in a short travel period, unobserved driver characteristics (e.g. psychological features) remain unchanged. Thus, unbiased associations between the observed driving characteristics and car-following safety can be established. It is noted that this premise may not stand when it comes to relatively longer travel during which drivers' psychological features may vary a lot. This demands future studies with further matching (e.g. reaction time) when long-distance trajectory data is available. Finally, 4,564 cases and 9,128 controls are identified from the US101 dataset. 2,157 cases and 4,314 controls are identified from the I-75 dataset.

Results

Pearson's correlation tests are conducted on both datasets before modelling to avoid the multiclonality issue. Test results are plotted in Figure 4 (a) and (b), respectively. Most variables are weakly correlated with others (i.e. coefficient less than 0.40) and only a few are moderately correlated with others (i.e. coefficient slightly greater than 0.40). Thus, the multiclonality issue will not raise (Kumari 2008; Shrestha 2020).

STATA 16 (Stata 2021) is used to estimate the conditional logistic model on the matched case–control data. The model estimation results are given in Table 3. The OR, confident

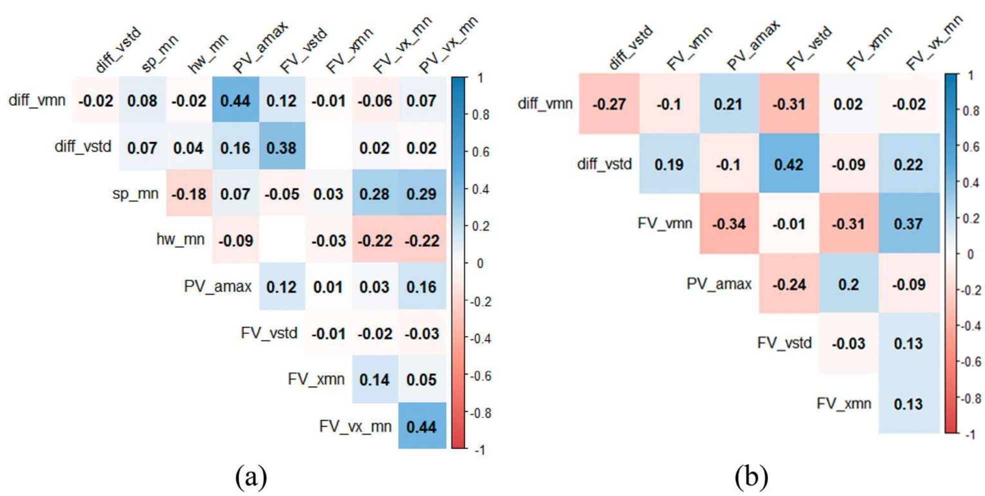

Figure 4. Pearson correlation tests; (a) US101 dataset; (b) I-75 dataset.

interval (CI) of OR, and standard error (SE) of OR are also presented. All variables are statistically significant at a CI of 95%.

Interaction risk factors

The longitudinal speed difference between the preceding vehicle and the following vehicle is found to be negatively associated with unsafe car-following. One unit increase in the longitudinal speed difference decreases the possibility of unsafe car-following to 0.209 times in the US101 model and 0.483 times in the I-75 model. This is intuitive since the faster the preceding vehicle travels, the smaller the possibility is that the following vehicle collides with it (Gu et al. 2019; Xu, Wang, and Liu 2013). This is also consistent with the data statistics presented in Table 2. The mean value of the longitudinal speed difference for all cases is only −0.45 m/s but 0.66 m/s for all controls in the US101 dataset, −1.29 m/s versus 0.31 m/s in the I-75 dataset. It is interesting to notice that the crash risk reduction magnitude is greater in the US101 model than in the I-75 model. This new finding suggests that it is safer if the following vehicle maintains a smaller speed than the preceding vehicle in more congested traffic.

The longitudinal speed difference standard deviation is positively associated with unsafe car-following. If the speed difference standard deviation increases by 1 m/s, the possibility of unsafe car-following increases to 5.493 times in the US101 model and 1.388 times in the I-75 model. This suggests that as the traffic becomes more oscillated, the crash risk greatly increases, which aligns with the existing literature (Pirdavani et al. 2015; Wang, Abdel-Aty, and Lee 2017; Wu, Abdel-Aty, and Lee 2018b). As shown in Table 2, the mean value of the longitudinal speed difference standard deviation for all cases is 0.88 m/s while it is only 0.47 for all control in the US101 dataset, 0.48 m/s versus 0.29 m/s in the I-75 dataset. Apart from consistent findings as in the existing literature, the impact magnitude of the longitudinal speed difference standard deviation is found to be greater in the US101 model than in the I-75 model. The possible reason is that the US101 traffic is more oscillated and congested, and thus it is more likely to expose to traffic crashes.

As expected, the average spacing is negatively associated with unsafe car-following. If the car-following spacing increases by 1 m, the possibility of unsafe car-following becomes 0.848 times in the US101 model. This highlights the significance of maintaining a long enough spacing to assure safety (Ni, Kang, and Andersen 2010; Yeung and Wong 2014). This is also reflected in Table 2 that the mean value of the average spacing for all cases is 14.71 m while it is 18.39 m for all controls in the US101 dataset. The average headway is also negatively associated with unsafe car-following in the US101 model. If the headway increases by 1 s, the possibility of unsafe car-following decreases to 0.517 times. However, such crash risk reduction is not observed in the I-75 model, i.e. the average spacing and average headway are not statistically significant. This indicates a new insight that the benefit of increasing the spacing and headway is more like to manifest in more congested traffic.

Individual vehicle risk factors

Longitudinal factors

The following vehicle longitudinal average speed is positively associated with unsafe car-following. If the average speed increases by 1 m/s, the corresponding possibility of unsafe car-following increases to 1.157 times in the I-75 model. The faster the following vehicle operates, the greater the crash risk is (Aarts and Van Schagen 2006). As shown in Table 2, the mean value of the following vehicle average longitudinal speed for all cases is 21.70 m/s yet it is only 15.53 m/s for all controls in the I-75 dataset. However, the safety effects of increasing the following vehicle longitudinal average speed are not statistically significant in the US101 model. This is probably because there is no significant difference between the following vehicle longitudinal average speed for all cases and controls, as indicated in Table 2, a mean of 8.57 m/s versus a mean of 7.39 m/s, in the congested traffic.

The preceding vehicle longitudinal maximum acceleration is negatively related to unsafe car-following. One unit increase in the maximum acceleration of the preceding vehicle is associated with 0.916 times unsafe car-following possibility in the US101 model and 0.834 times in the I-75 model. With a greater maximum acceleration, the preceding vehicle can travel faster, and consequently the crash risk decreases (Gu et al. 2019; Xu, Wang, and Liu 2013). Also, a new finding is yielded that the crash risk reduction magnitude is greater in the I-75 model. Since the I-75 traffic is less congested, a greater maximum acceleration is expected to result in an even greater preceding vehicle longitudinal speed than the US101 traffic. Consequently, crash risks are much reduced.

The following vehicle longitudinal speed standard deviation is positively associated with unsafe car-following. One unit increase contributes to 1.925 times unsafe car-following probability in the US101 model and 2.264 times in the I-75 model. As the following vehicle becomes less longitudinally stable, the crash risks increase (Pirdavani et al. 2015). The magnitude of the threatening effects is greater in the I-75 model than in the US101 model, which is probably accounted for by the apparently greater speed of the I-75 traffic. This new finding illustrates the importance of keeping longitudinally stable, i.e. avoiding frequent acceleration and/or deceleration, in assuring car-following safety especially when the operating speed is relatively great.

Lateral factors

When considering the individual vehicle lateral risk factors, more interesting and novel findings are drawn as follows.

First, the following vehicle average deviation from the lane centre is found to increase the possibility of unsafe car-following. If the average deviation increases by 1 m, the unsafe car-following possibility is expected to increase to 2.712 times in the US101 model and 1.627 times in the I-75 model. Similarly, the following vehicle average lateral speed is also positively associated with the unsafe car-following possibility. If the lateral speed increases by 0.1 m/s, the corresponding possibility of unsafe car-following increases to 1.566 times in the US101 model and 1.226 times in the I-75 model. Both the above factors are indicators of unstable lane-keeping behaviour. When the following vehicles are more laterally unstable, it is of great possibility that the drivers are distracted. This would contribute to higher crash risks. It is also noticed that the threatening effects of unstable lane-keeping behaviour of the following vehicle on unsafe car-following are greater in the US101 model. The crash risks are much greater when vehicles are more laterally unstable in more congested traffic.

The preceding vehicle average lateral speed is also positively related to the unsafe car-following possibility. If the average lateral speed increases by 0.1 m/s, the associated unsafe car-following possibility increases to 1.764 times in the US101 model. As the preceding vehicle becomes more laterally unstable, the crash risks are expected to climb. However, this variable is not significant in the I-75 model. The laterally unstable behaviour of the preceding vehicles is less likely to increase the crash risks when the traffic is less congested and the car-following spacing is longer.

It should be noted that the above explanation does not contradict the discussion regarding the threatening effects of the preceding vehicle longitudinal speed standard deviation, i.e. longitudinally unstable behaviour indicator. Because the impacts of the preceding vehicle longitudinal speed standard deviation on the unsafe car-following probability are direct while the impacts of the preceding vehicle average lateral speed are indirect that are related to unobserved driver heterogeneity which cannot be controlled by matching vehicle IDs, e.g. whether a driver is distracted or not. It may be possible to control such driver heterogeneity by additionally matching the drivers' reaction time based on the premise that a driver's reaction time is longer when he/she is distracted. Given the short car-following duration of each preceding-following vehicle pair in the two datasets, i.e. mean values around 36 s, further matching the drivers' reaction time would result in an extremely small sample size. Thus, we list this as a future research direction when long-duration car-following trajectory data is available. It is interesting to investigate the car-following safety impacts of the lateral characteristics after the further matching.

Conclusion

This study adopts a matched case–control approach to quantify the safety effects of observed driving characteristics on car-following safety. The confounding effects of unobserved driver heterogeneity are controlled by matching preceding vehicle ID and following vehicle ID. The NGSIM US101 and HIGHSIM I-75 datasets are processed and matched. The conditional logistic regression model is estimated on the matched data to estimate the safety effects of both longitudinal (e.g. car-following) and lateral (e.g. lane-keeping) driving characteristics.

Some findings are found to be consistent with the existing studies. The longitudinal average speed difference, average spacing, average headway, and preceding vehicle longitudinal maximum acceleration are negatively related to car-following crash risks. The longitudinal speed difference standard deviation, following vehicle longitudinal average speed, and following vehicle longitudinal speed standard deviation are positively associated with car-following crash risks.

Several key conclusions and new insights are drawn. (1). The following vehicle average deviation from the lane centre, following vehicle average lateral speed, and preceding vehicle average lateral speed are found to increase the possibility of unsafe car-following, indicating that unstable lateral movements contribute to higher crash risks. Therefore, a stable and efficient lane-keeping technology is essential in autonomous vehicle development. Regarding human drivers, distracted driving (e.g. texting, eating, and drinking while driving) that causes unstable lateral movements should be avoided. Advanced driver assistance systems with an unstable lateral movement warning function should be developed. Unlike the widely deployed lane marking detection that only warns drivers when the vehicle touches the lane marking, this function can be configured to alert drivers when the vehicle has a lateral speed greater than a threshold. (2). When traffic is more congested, there is less room on the roadway surface that can be used to maneuver a risky car-following scenario, i.e. the traffic system is fragile and even a small change can easily increase crash risks.

Therefore, the following vehicle is suggested to maintain a greater speed difference, spacing, and headway from the preceding vehicle in more congested traffic. Besides, vehicles are encouraged to keep laterally as stable as possible to reduce more crash risks in more congested traffic.

The above findings enable a more comprehensive and accurate understanding of car-following risk factors with unobserved driver-specific characteristics controlled. These findings are valuable in enhancing traffic safety management and in helping develop a real-time car-following safety warning system for automated vehicles to facilitate their safe operations.

Yet, limitations exist in this study due to the short car-following duration of each vehicle pair in the two datasets. In the future, it is worthy to further control driver heterogeneity by additionally matching the drivers' reaction time when long-duration car-following trajectory data is available based on the premise that even the same driver could respond differently. Further, this study does not classify cases into different risk levels. This is a standard treatment of matched case–control study but may lose the information of the possible crash risks. In a future study, the car-following safety indicator can be formulated as an ordinal variable instead of a binary variable to address this limitation. It would be valuable to investigate the relationship between actual crashes and risk factors in the future when relevant data is available. The car-following risk model spatial transferability can be tested on car-following data collected at the same time (probably the same year) but at a different location, when available. The model temporal transferability can be tested on car-following data collected at the same location but at a different time, when available. Last, although some seeming risk factors (e.g. vehicle type and lane number) are not significant in the investigated two datasets, their impacts on car-following risks shall be tested when new datasets are available.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by U.S. National Science Foundation [grant number CMMI #1932452].

ORCID

Qianwen Li http://orcid.org/0000-0002-5901-8702
Handong Yao http://orcid.org/0000-0002-2778-084X
Xiaopeng Li http://orcid.org/0000-0002-5264-3775

References

Aarts, L., and I. Van Schagen. 2006. "Driving Speed and the Risk of Road Crashes: A Review." *Accident Analysis & Prevention* 38 (2): 215–224.

AASHTO, A. 2001. *Policy on Geometric Design of Highways and Streets*. Washington, DC: The American Association of State Highway and Transportation Officials.

Aksan, N., L. Sager, S. Hacker, R. Marini, J. Dawson, S. Anderson, and M. Rizzo. 2016. "Forward Collision Warning: Clues to Optimal Timing of Advisory Warnings." *SAE International Journal of Transportation Safety* 4 (1). doi:10.4271/2016-01-1439.

Behbahani, H., N. Nadimi, and S. S. Naseralavi. 2015. "New Time-Based Surrogate Safety Measure to Assess Crash Risk in car-Following Scenarios." *Transportation Letters* 7 (4): 229–238. doi:10.1179/1942787514Y.0000000051.

Blanco, M., J. Atwood, S. M. Russell, T. Trimble, J. A. McClafferty, and M. A. Perez. 2016. *Automated Vehicle Crash Rate Comparison Using Naturalistic Data*. Virginia Tech Transportation Institute.

Breslow, N. E. 1996. "Statistics in Epidemiology: The Case-Control Study." *Journal of the American Statistical Association* 91 (433): 14–28.

Breslow, N. E., N. E. Day, K. T. Halvorsen, R. L. Prentice, and C. Sabai. 1978. "Estimation of Multiple Relative Risk Functions in Matched Case-Control Studies." *American Journal of Epidemiology* 108 (4): 299–307.

Chen, Q., R. Gu, H. Huang, J. Lee, X. Zhai, and Y. Li. 2021a. "Using Vehicular Trajectory Data to Explore Risky Factors and Unobserved Heterogeneity During Lane-Changing." *Accident Analysis & Prevention* 151, 105871. doi:10.1016/j.aap.2020.105871.

Chen, Q., H. Huang, Y. Li, J. Lee, K. Long, R. Gu, and X. Zhai. 2021b. "Modeling Accident Risks in Different Lane-Changing Behavioral Patterns." *Analytic Methods in Accident Research* 30. doi:10.1016/j.amar.2021.100159.

Edie, L. C. 1963. "Discussion of Traffic Stream Measurements and Definitions".International Symposium on the Theory of Traffic Flow and Transportation. London, England.

FHWA. 2008. The Next Generation Simulation (NGSIM) [Online] [WWW Document]. http://www.ngsim.fhwa.dot.gov/.

Gjerde, H., P. T. Normann, A. S. Christophersen, S. O. Samuelsen, and J. Mørland. 2011. "Alcohol, Psychoactive Drugs and Fatal Road Traffic Accidents in Norway: A Case–Control Study." *Accident Analysis & Prevention* 43 (3): 1197–1203.

Gross, F. 2013. "Case–Control Analysis in Highway Safety: Accounting for Sites with Multiple Crashes." *Accident Analysis & Prevention* 61: 87–96.

Gu, X., M. Abdel-Aty, Q. Xiang, Q. Cai, and J. Yuan. 2019. "Utilizing UAV Video Data for in-Depth Analysis of Drivers' Crash Risk at Interchange Merging Areas." *Accident Analysis & Prevention* 123: 159–169. doi:10.1016/j.aap.2018.11.010.

Híjar, M., C. Carrillo, M. Flores, R. Anaya, and V. Lopez. 2000. "Risk Factors in Highway Traffic Accidents: A Case Control Study." *Accident Analysis & Prevention* 32 (5): 703–709.

Huertas-Leyva, P., M. Nugent, G. Savino, M. Pierini, N. Baldanzini, and S. Rosalie. 2019. "Emergency Braking Performance of Motorcycle Riders: Skill Identification in a Real-Life Perception-Action Task Designed for Training Purposes." *Transportation Research Part F: Traffic Psychology and Behaviour* 63: 93–107.

Ingram, D., K. Sanders, M. Kolybaba, and D. Lopez. 1997. "Case-control Study of Phyto-Oestrogens and Breast Cancer." *The Lancet* 350 (9083): 990–994.

James, R. M., C. Melson, J. Hu, and J. Bared. 2019. "Characterizing the Impact of Production Adaptive Cruise Control on Traffic Flow: An Investigation." *Transportmetrica B: Transport Dynamics* 7 (1): 992–1012.

Jamson, A. H., F. C. H. Lai, and O. M. J. Carsten. 2008. "Potential Benefits of an Adaptive Forward Collision Warning System." *Transportation Research Part C: Emerging Technologies* 16 (4): 471–484.

Khoda Bakhshi, A., and M. M. Ahmed. 2021. "Utilizing Black-box Visualization Tools to Interpret Non-parametric Real-Time Risk Assessment Models." *Transportmetrica A: Transport Science* 17 (4): 739–765. doi:10.1080/23249935.2020.1810169.

Knoop, V. L., S. P. Hoogendoorn, Y. Shiomi, and C. Buisson. 2012. "Quantifying the Number of Lane Changes in Traffic." *Transportation Research Record: Journal of the Transportation Research Board* 2278 (1): 31–41.

Koopman, P., and M. Wagner. 2017. "Autonomous Vehicle Safety: An Interdisciplinary Challenge." *IEEE Intelligent Transportation Systems Magazine* 9 (1): 90–96.

Kumari, S. S. 2008. "Multicollinearity: Estimation and Elimination." *Journal of Contemporary Research in Management* 3 (1): 87–95.

Kweon, Y.-J. 2011. "Development of Crash Prediction Models with Individual Vehicular Data." *Transportation Research Part C: Emerging Technologies* 19 (6): 1353–1363.

Laval, J. A. 2011. "Hysteresis in Traffic Flow Revisited: An Improved Measurement Method." *Transportation Research Part B: Methodological* 45 (2): 385–391.

Li, Y., D. Wu, J. Lee, M. Yang, and Y. Shi. 2020. "Analysis of the Transition Condition of Rear-End Collisions Using Time-to-Collision Index and Vehicle Trajectory Data." *Accident Analysis & Prevention* 144. doi:10.1016/j.aap.2020.105676.

Lu, X. Y., P. Varaiya, and R. Horowitz. 2009. "Fundamental Diagram Modelling and Analysis Based NGSIM Data." *IFAC Proceedings* 42 (15): 367–374. doi:10.3182/20090902-3-US-2007.0068.

Ni, R., J. J. Kang, and G. J. Andersen. 2010. "Age-related Declines in car Following Performance Under Simulated Fog Conditions." *Accident Analysis & Prevention* 42 (3): 818–826.

Oh, C., and T. Kim. 2010. "Estimation of Rear-end Crash Potential Using Vehicle Trajectory Data." *Accident Analysis & Prevention* 42 (6): 1888–1893. doi:10.1016/j.aap.2010.05.009.

Park, H., C. Oh, J. Moon, and S. Kim. 2018. "Development of a Lane Change Risk Index Using Vehicle Trajectory Data." *Accident Analysis & Prevention* 110: 1–8. doi:10.1016/j.aap.2017.10.015.

Peng, L., N. Lyu, and C. Wu. 2020. "A Road Traffic Crash Risk Assessment Method Using Vehilce Trajectory Data and Surrogate Safety Measures." *CICTP Tan* 2011: 3144–3155.

Pirdavani, A., E. De Pauw, T. Brijs, S. Daniels, M. Magis, T. Bellemans, and G. Wets. 2015. "Application of a Rule-Based Approach in Real-Time Crash Risk Prediction Model Development Using Loop Detector Data." *Traffic Injury Prevention* 16 (8): 786–791.

Przybyla, J., J. Taylor, J. Jupe, and X. Zhou. 2015. "Estimating Risk Effects of Driving Distraction: A Dynamic Errorable car-Following Model." *Transportation Research Part C: Emerging Technologies* 50: 117–129. doi:10.1016/j.trc.2014.07.013.

Schlesselman, J. J. 1982. *Case-Control Studies: Design, Conduct, Analysis*. New York: Oxford University Press.

Shi, X., Y. D. Wong, M. Z. F. Li, and C. Chai. 2018. "Key Risk Indicators for Accident Assessment Conditioned on Pre-crash Vehicle Trajectory." *Accident Analysis & Prevention* 117: 346–356. doi:10.1016/j.aap.2018.05.007.

Shi, X., D. Zhao, H. Yao, X. Li, D. K. Hale, and A. Ghiasi. 2021. "Video-based Trajectory Extraction with Deep Learning for High-Granularity Highway Simulation (HIGH-SIM)." *Communications in Transportation Research* 1: 100014.

Shrestha, N. 2020. "Detecting Multicollinearity in Regression Analysis." *American Journal of Applied Mathematics and Statistics* 8 (2): 39–42.

Song, Y., M. V. Chitturi, and D. A. Noyce. 2021. "Automated Vehicle Crash Sequences: Patterns and Potential Uses in Safety Testing." *Accident Analysis & Prevention* 153: 106017.

Stata.com. 2021. clogit — Conditional (fixed-effects) logistic regression [WWW Document].

statista. 2017. Projected Number of Autonomous Vehicles in Operation in the United States in 2025 and 2030 [WWW Document], 201.

Wang, L., M. Abdel-Aty, and J. Lee. 2017. "Safety Analytics for Integrating Crash Frequency and Real-Time Risk Modeling for Expressways." *Accident Analysis & Prevention* 104: 58–64.

Wey, S. B., M. Mori, M. A. Pfaller, R. F. Woolson, and R. P. Wenzel. 1989. "Risk Factors for Hospital-Acquired Candidemia: A Matched Case-Control Study." *Archives of Internal Medicine* 149 (10): 2349–2353.

WHO. 2021. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries [WWW Document].

Woodward, M. 2013. *Epidemiology: Study Design and Data Analysis*. New York: CRC Press.

Wu, Y., M. Abdel-Aty, Q. Cai, J. Lee, and J. Park. 2018a. "Developing an Algorithm to Assess the Rear-end Collision Risk Under fog Conditions Using Real-Time Data." *Transportation Research Part C: Emerging Technologies* 87: 11–25.

Wu, Y., M. Abdel-Aty, and J. Lee. 2018b. "Crash Risk Analysis During fog Conditions Using Real-Time Traffic Data." *Accident Analysis & Prevention* 114: 4–11.

Xie, K., K. Ozbay, H. Yang, and C. Li. 2019. "Mining Automatically Extracted Vehicle Trajectory Data for Proactive Safety Analytics." *Transportation Research Part C: Emerging Technologies* 106: 61–72. doi:10.1016/j.trc.2019.07.004.

Xu, C., W. Wang, and P. Liu. 2013. "Identifying Crash-Prone Traffic Conditions Under Different Weather on Freeways." *Journal of Safety Research* 46: 135–144.

Xue, Q., K. Wang, J. J. Lu, and Y. Liu. 2019. "Rapid Driving Style Recognition in Car-Following Using Machine Learning and Vehicle Trajectory Data." *Journal of Advanced Transportation* 2019: 1. doi:10.1155/2019/9085238.

Yeung, J. S., and Y. D. Wong. 2014. "The Effect of Road Tunnel Environment on car Following Behaviour." *Accident Analysis & Prevention* 70: 100–109.

Yu, R., L. Han, and H. Zhang. 2021. "Trajectory Data Based Freeway High-Risk Events Prediction and its Influencing Factors Analyses." *Accident Analysis & Prevention* 154: 106085. doi:10.1016/j.aap.2021.106085.

Zhang, N., G. M. Dong, and H. P. Du. 2008. "Investigation Into Untripped Rollover of Light Vehicles in the Modified Fishhook and the Sine Maneuvers. Part I: Vehicle Modelling, Roll and Yaw Instability." *Vehicle System Dynamics* 46 (4): 271–293. doi:10.1080/00423110701344752.

Zhang, J., Y. Wang, and G. Lu. 2019. "Impact of Heterogeneity of Car-Following Behavior on Rear-end Crash Risk." *Accident Analysis & Prevention* 125: 275–289. doi:10.1016/j.aap.2019.02.018.

Zhao, P., and C. Lee. 2018. "Assessing Rear-end Collision Risk of Cars and Heavy Vehicles on Freeways Using a Surrogate Safety Measure." *Accident Analysis & Prevention* 113: 149–158. doi:10.1016/j.aap.2018.01.033.

Appendix

This appendix presents the model results when TTC is adopted to measure the car-following safety. The threshold of differentiating risky car-following from safe car-following is set as 4 s by referring to the existing literature (Aksan et al. 2016). Within the car-following segment, the TTC value is calculated for each time step. If all TTC values are greater than 4 s, the segment is labelled as a control (i.e. safe). If at least one TTC value is smaller than 4 s, the segment is labelled as a case (i.e. risky).

The detailed model fitting results are provided in Table A1. It is observed that the effects of risk factors on car-following safety when TTC is used with a threshold of 4 s are consistent with those when SDI is used (Table 3) qualitatively. Specifically, the longitudinal average speed difference and average spacing are negatively related to car-following crash risks. The longitudinal speed difference standard deviation, following vehicle longitudinal speed standard deviation, and following vehicle average lateral speed are positively related to car-following crash risks. The comparison results between the NGSIM US101 and HIGHSIM I-75 when TTC is used with a threshold of 4 s are also consistent with those when SDI is used (Table 3) qualitatively. It is safer when the following vehicle maintains a smaller speed than the preceding vehicle in more congested traffic and when the following vehicle stays longitudinally stable especially when the overall traffic speed is great.

Some significant risk factors in Table 3 are not significant anymore when TTC is used as the safety measurement. Yet, there are also new significant factors observed. The following vehicle longitudinal maximum acceleration and preceding vehicle longitudinal speed standard deviation are found to be positively related to car-following risks.

It should be noted that different SSMs (e.g. SDI and TTC) and even the same SSM with different threshold values differentiating risky and safe situations (e.g. TTC $= 4$ s and TTC $= 7$ s) are expected to yield different model estimation results in terms of specific model parameter values. This is intuitive because the values of the response variable (i.e. either safe [0] or risky [1]) have changed. However, the qualitative impact of each risk factor remains the same, as discussed above.

Table A1. Fitted conditional logistic model results (TTC = 4 s).

	US101						I-75					
Variable	Coef.	***z***	*p*-value	OR	95% CI of OR	SE of OR	Coef.	***z***	*p*-value	OR	95% CI of OR	SE of OR
Interaction risk factors												
diff_vmn (m/s)		−21.60	0.000	0.103	[0.087, 0.123]	0.011		−2.10	0.036	0.834	[0.704, 0.988]	0.072
diff_vstd (m/s)								3.67	0.000	3.189	[1.717, 5.923]	1.008
sp_mn (m)		−22.28	0.000	0.596	[0.574, 0.619]	0.014						
Individual vehicle risk factors-longitudinal factors												
FV_amax[1] (m/s^2)		1.70	0.008	1.103	[1.002, 1.193]	0.059						
PV_vstd[2] (m/s)		12.14	0.000	4.648	[3.775, 5.725]	0.588						
FV_vstd (m/s)		3.79	0.000	1.674	[1.339, 2.093]	0.227		2.33	0.020	2.088	[1.123, 3.884]	0.661
1Individual vehicle risk factors-lateral factors												
FV_vx_mn (0.1 m/s)								2.24	0.025	2.556	[1.125, 5.811]	1.071
Model Statistics												
Number of observations				6,378						204		
Log-likelihood				−633.989						−28.376		
Pseudo R^2				0.729						0.620		

1: FV_amax = following vehicle longitudinal maximum acceleration.
2: PV_vstd = preceding vehicle longitudinal speed standard deviation.
All other variables have been introduced in Table 2.