Estimating Rear-End Accident Probabilities with Different Driving Tendencies at Signalized Intersections in China

Weijie Wang,1 Yingshuai Li,1 Jian Lu,2,3,4 Yaping Li,2,3,4 and Qian Wan5

1School of Transportation, Nanjing Tech University, Nanjing 210009, China
2Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing 210096, China
3Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Si Pai Lou #2, Nanjing 210096, China
4School of Transportation, Southeast University, Si Pai Lou #2, Nanjing 210096, China
5Hualan Design & Consulting Group, Hua Dong Lu #39, Nanning 530011, China

Correspondence should be addressed to Yingshuai Li; tianwang0726@126.com

Received 16 July 2018; Revised 30 November 2018; Accepted 6 January 2019; Published 20 January 2019

Copyright © 2019 Weijie Wang 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.

Rear-end accidents are the most common accident type at signalized intersections because of the different driving tendencies in the dilemma zone (DZ), where drivers are faced with indecisiveness of making "stop or go" decisions at yellow onset. In various researches, the number of vehicles in the DZ has been used as a safety indicator—the more the vehicles in the DZ, the higher the probability of rear-end accidents. However, the DZ-associated rear-end accident potential varies depending on drivers’ driving tendencies and the situations (position and speed) at the yellow onset. This study’s primary objective is to explore how the driving tendency impacts the DZ distribution and the probability of rear-end accidents. To achieve this, three types of driving tendencies were classified using K-means clustering analysis based on driving variables. Further, the boundary of the DZ is determined by logistic regression model of drivers’ stop/go decision. Then, we proposed the conditional probability model of rear-end accidents and developed a Monte Carlo simulation framework to calculate the model. The results indicate that the rear-end accident probability is dependent on the driving tendency even at the same position with the same speed in the DZ. The aggressive type has the highest risk probability followed by conservative and then the normal types. The quantitative results of the study can provide the basis for rear-end accident assessments.

1. Introduction

Rear-end accidents are common at signalized intersections in China. Besides driver errors, indecisiveness in the dilemma zone (DZ) at signalized intersections is a leading cause of such accidents [1]. To assist drivers with their decision making during the critical phase transition interval, green signal countdown devices (GSCDs) display the remaining seconds of the current signal status through synchronizing with the traffic signal controller. This provides drivers with information on the current phase’s termination in advance. GSCD installation may improve the utilization of the amber time and the capacity of intersection approach road. However, it may also easily evoke risky car-following behavior that results in a much higher incidence of rear-end accidents in the DZ during the flashing green and yellow intervals.

Although some negative effects of the GSCDs installation on intersection safety have been observed by some researchers [2, 3] and the view of scholars on this issue is inconsistent [4–8], most urban traffic managers in China are still inclined to install GSCDs. Such phenomenon is mainly due to China’s unique driving culture, local enforcement activities, and public perceptions. To reduce the probability of rear-end accidents, a new regulation of “solid yellow” was implemented at the beginning of 2013. The regulation stipulates that “drivers who had already entered the intersection could pass through it, but those who had not should stop”. When the yellow light is displayed, drivers who crossed stop
line “violated yellow” would have six points (half of their 12-point limit) deducted if caught.

An issue is created when intersections do not have GSCDs (NGSCDs), as drivers do not know when the green light changes to the yellow light. When the lights change, if a driver is close to the stop line, they have to stop abruptly, which leads to a higher risk of rear-end accidents. Finally, the regulation and enforcement were forced to revive again after being implemented for a week that drivers should stop if they can stop safely when the signal turns to yellow; however, those who cross the stop line on the yellow light will not be punished [9].

The “solid yellow” regulation was only implemented for a week; however, the discussion it triggered is not over. The big issue of NGSCDs during the regulation’s implementation has made the public and urban traffic managers aware of the advantages of GSCDs, causing GSCDs to be increasingly used in many cities of China, such as Beijing, Shanghai, Guangzhou, Nanjing, Shijiazhuang, and Harbin [10].

We believe that some of the negative impact of GSCDs on the DZ may partially come from the effect of characteristics of drivers. Since the last decade, a growing number of Chinese people have obtained their own cars and licenses, which has directly resulted in an increase in the number of novice drivers in the population [11]. This has led to a growing difference in the various driver attributes (including age, gender, driving experience, and educational background), which adds to the definition of “heterogeneous traffic”. Heterogeneous traffic not only makes the distribution of DZ more discrete but also increases the uncertainty of drivers’ response to GSCDs, especially to the phase transition interval. Therefore, the risk of rear-end collision at signalized intersection in China has significantly increased.

DZ can be divided into two categories: type I and type II [12]. Type I DZ was initially proposed by Gazis, Herman, and Maradudin [13] (GHM), who defined it as an area in which drivers can neither stop comfortably nor cross safely before the signal turns red.

As shown in Figure 1, when \( X_c > X_0 \), the area between \( X_c \) and \( X_0 \) is the dilemma zone \( (X_d = X_c - X_0) \); however, when \( X_c < X_0 \), the zone is termed the option zone. In the option zone, drivers can either pass the intersection or slow down and stop at the stop line while the yellow signal is on or they can go. If the driver hesitates, it will lead to the emergence of the DZ. In fact, the boundaries of type I DZ are not fixed, and it will change depending on the driver’s speed on the yellow light’s onset [15–17]. In addition, some drivers use the yellow interval as an extension of the green light phase [18]. Roni Factor et al. [19] observed that Israeli drivers exhibit great variance in their reaction to the flashing green. May [20] concluded that some vehicles accelerated/decelerated heavily to escape the DZ. Liu et al. [21] observed that driver behaviors were considerably different from the theoretical assumptions. These studies imply that vehicles’ approaching speeds and drivers’ behaviors have a wide distribution and variations. A lack of capability to tackle the randomness in the driving behavior is the main disadvantage of the GHM model [1].

If the concept of type I DZ is from the view of stopping and yellow light passing distances, then the concept of type II DZ is from the drivers’ choice of stopping, which Zegeer [22] defined as “an area where more than 10 percent and less than 90 percent of drivers would decide to stop when the yellow is presented”. According to this concept, the boundaries of type II DZ are determined from “distance from stop line” and “time to intersection” (TTI) [23–25], which change with the driving tendency. Fu [26] indicated in his doctoral dissertation that when there is no pedestrian on the crosswalk, type I DZ is included in type II DZ when the yellow signal is at three seconds or four seconds.

Estimating accident risk probabilities of the DZ is a key issue in the DZ protection systems. The traditional approach is to take the number of vehicles in the DZ as a risk measurement indicator of the DZ at yellow onset. This approach is based on the assumption that the number of vehicles in the DZ is equivalent to the corresponding accident risk probabilities of the DZ, regardless of where the vehicle is located. However, the accident risk probabilities of the DZ will change based on factors such as the location and speeds of vehicles, the drivers’ behavior, and the power and braking performance of the vehicles. Some scholars have been exploring this field from the perspective of transition interval phase [27], with or without DSCDs [28], driver behavior [21, 29], and the relationship among them [30].

Past studies made significant contributions to understanding the impact of countdown timers, driver behavior, the phase transition interval (usually referred to as the yellow signal),
and the boundaries of the DZ on the accident risk probabilities of the DZ. However, the following issues remain to be addressed:

(1) Previous studies only investigated the impact of GSCDs or NGSCDs concerning the stop/go decisions of drivers but neglected the different driving maneuvers (e.g., aggressive acceleration, abrupt stop, and headway) of different driving tendencies (aggressive, normal, and conservative) under the condition of GSCDs. As we know, with an increasing number of Chinese people obtaining their own cars and licenses, the diversity of drivers is also increasing. Differences in driving tendency add meaning to “heterogeneous traffic”, which increases the complexity of driving behavior in the DZ and results in serious rear-end accidents.

(2) Although GSCDs have been installed at many urban intersections in China, with the significantly diverse driving behavior and China’s fast-growing roadway system, more field studies are needed to enrich the dataset for developing regulations.

This paper is aimed at developing a model to quantify the risk probabilities of rear-end accidents under different driving tendencies during the phase transition interval. More specifically, the research objective includes the following tasks: (1) classifying drivers into three types (aggressive, normal, and conservative) based on their responses to phase transition interval; (2) proposing a comparison of the driving parameters (speeds, acceleration/deceleration, and headway) of the three driver types and fitting equation; (3) establishing the driver’s stop/go binary logistic probability model; (4) proposing a conditional probability model of rear-end accidents; (5) simulating the conditional probability model by Monte Carlo simulation. The research results can provide the basis for rear-end accident assessment.

2. Data Collection

2.1. Study Sites. We chose two urban signalized intersections as the experimental observation points, in Nanjing, China. The intersections were chosen because they are representative and have the same traffic flow patterns (97 percent of passenger cars and 3 percent of buses), the same signal phase (follow the same sequence of “steady green-flashing green (FG; 3s)-yellow (Y; 3s)-red”, and each intersection is not set all-red time, and the yellow light is the clearance time), and the same speed limit value (60km/h). For each intersection, one camera was set up on top of a roadside building to cover a studied approach road. The first studied approach is an eastbound approach along ZhongShan East Avenue (~120 m); the second studied approach is an eastbound approach along HeXi Avenue (~280 m). The video for each intersection approach was recorded from 8:00 a.m. to noon and 1:00 p.m. to 5:00 p.m. For each studied approach, over 150 hours of recorded video data were collected. The video recording process is shown in Figures 2(a) and 2(b).
Table 1: Statistical results of driving variable parameters of different driving tendencies.

<table>
<thead>
<tr>
<th>Driving tendencies</th>
<th>Variables</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aggressive</strong></td>
<td>Distance from stop line (m)</td>
<td>220</td>
<td>24.34</td>
<td>69.79</td>
<td>41.1618</td>
<td>7.4732</td>
</tr>
<tr>
<td></td>
<td>Speed (m/s)</td>
<td>220</td>
<td>10.11</td>
<td>18.48</td>
<td>13.3690</td>
<td>1.7904</td>
</tr>
<tr>
<td></td>
<td>Acceleration/Deceleration (m/s²)</td>
<td>220</td>
<td>-4.16</td>
<td>1.40</td>
<td>-1.0895</td>
<td>1.66172</td>
</tr>
<tr>
<td></td>
<td>Entry time¹ (s)</td>
<td>131</td>
<td>2.64</td>
<td>5.22</td>
<td>3.5204</td>
<td>.57218</td>
</tr>
<tr>
<td></td>
<td>Headway (s)</td>
<td>213</td>
<td>.67</td>
<td>2.99</td>
<td>1.7348</td>
<td>.58343</td>
</tr>
<tr>
<td></td>
<td>Stop time² (s)</td>
<td>89</td>
<td>4.10</td>
<td>6.01</td>
<td>4.8156</td>
<td>.43181</td>
</tr>
<tr>
<td></td>
<td>Reaction time (s)</td>
<td>220</td>
<td>.32</td>
<td>1.31</td>
<td>.8276</td>
<td>.15027</td>
</tr>
<tr>
<td><strong>Normal</strong></td>
<td>Distance from stop line (m)</td>
<td>352</td>
<td>19.97</td>
<td>67.49</td>
<td>36.3597</td>
<td>6.62826</td>
</tr>
<tr>
<td></td>
<td>Speed (m/s)</td>
<td>352</td>
<td>5.27</td>
<td>18.53</td>
<td>12.6959</td>
<td>1.63167</td>
</tr>
<tr>
<td></td>
<td>Acceleration/Deceleration (m/s²)</td>
<td>352</td>
<td>-2.64</td>
<td>4.24</td>
<td>-.5775</td>
<td>2.06079</td>
</tr>
<tr>
<td></td>
<td>Entry time¹ (s)</td>
<td>150</td>
<td>1.63</td>
<td>3.19</td>
<td>2.3965</td>
<td>.33242</td>
</tr>
<tr>
<td></td>
<td>Headway (s)</td>
<td>296</td>
<td>.65</td>
<td>3.73</td>
<td>1.9953</td>
<td>.61555</td>
</tr>
<tr>
<td></td>
<td>Stop time² (s)</td>
<td>202</td>
<td>4.27</td>
<td>8.01</td>
<td>5.8085</td>
<td>.69168</td>
</tr>
<tr>
<td></td>
<td>Reaction time (s)</td>
<td>352</td>
<td>.35</td>
<td>1.95</td>
<td>1.0744</td>
<td>.25885</td>
</tr>
<tr>
<td><strong>Conservative</strong></td>
<td>Distance from stop line (m)</td>
<td>224</td>
<td>15.89</td>
<td>76.42</td>
<td>38.9934</td>
<td>11.93947</td>
</tr>
<tr>
<td></td>
<td>Speed (m/s)</td>
<td>224</td>
<td>3.98</td>
<td>17.13</td>
<td>11.7563</td>
<td>1.76093</td>
</tr>
<tr>
<td></td>
<td>Acceleration/Deceleration (m/s²)</td>
<td>224</td>
<td>-2.37</td>
<td>2.45</td>
<td>-.6588</td>
<td>1.02378</td>
</tr>
<tr>
<td></td>
<td>Entry time¹ (s)</td>
<td>116</td>
<td>1.62</td>
<td>3.28</td>
<td>2.4757</td>
<td>.29486</td>
</tr>
<tr>
<td></td>
<td>Headway (s)</td>
<td>170</td>
<td>.98</td>
<td>3.99</td>
<td>2.0137</td>
<td>.59860</td>
</tr>
<tr>
<td></td>
<td>Stop time² (s)</td>
<td>108</td>
<td>5.59</td>
<td>12.10</td>
<td>8.5019</td>
<td>1.43955</td>
</tr>
<tr>
<td></td>
<td>Reaction time (s)</td>
<td>224</td>
<td>.44</td>
<td>2.46</td>
<td>1.2434</td>
<td>.41000</td>
</tr>
</tbody>
</table>

Note: (1) Entry time refers to the time from flashing green onset to the time when the vehicle has just entered the intersection. (2) Stop time refers to the time from flashing green onset to the time when the vehicle stops at the stop line.

2.2. Trajectory Data Extraction. The extraction method of trajectory data involved two steps. Firstly, the Tracker Video Analysis and Modeling Tool were used to extract the vehicles' coordinates in the video. Secondly, the imaging principle was used to match the coordinates in the video with the location in the real world.

Tracker [31] is a free video analysis and modeling tool built on the Open Source Physics (OSP) Java framework. It has manual and automated object tracking with position, velocity, and acceleration overlays and data. In this paper, Tracker was used for extracting and tracking the vehicles' coordinates. First, we need to put the video into Tracker and then build a reference frame. Track points were set up on the right corner of the front windscreen and the sampling frequency was set to be five frames per second. As shown in Figure 2(c), Tracker can generate real-time coordinates and space-time diagrams of subject vehicles while producing three columns of data—time (t), x-coordinate, and y-coordinate—to record the coordinates of the subject vehicles. Second, the imaging principle was used to match the coordinates in the video to the real world location. We selected eight feature points (Figure 2(d)), where B, C, E, and G are the coordinate points and A, D, F, and H are the error-checking points.

In this study, we include only two-vehicle accidents or the collision of the first two vehicles in an accident involving more than two vehicles. Approximately 796 couples’ trajectory data was extracted. Then, the driving parameters were derived, such as the speed, the acceleration/deceleration, and the headways.

3. Data Analysis

3.1. Driving Tendency Classification. In previous studies, driving tendency classification is completed according to the comparison of drivers’ actual stop/go decisions with the expected stop/go decision [9, 21] which is based on the theoretical calculating value. For example, Ni [9] defined aggressive drivers as drivers who aggressively pass through the intersection during the yellow phase even though the distance from the stop line \((D)\) is longer than \(X_0\) (the theoretical calculating value of the maximum distance a vehicle can pass the intersection at the highest driving efficiency before the red signal is shown). With this definition, one can exclude aggressive drivers who stop at the stop line using aggressive acceleration or abrupt braking during their driving. In addition, from Liu [21], it can be seen that there is a big difference between the theoretical values and field measured values. Therefore, we believe the data-driven method may be more realistic than the mechanism analysis method.

Therefore, in this paper, the K-means clustering analysis method was adopted to classify the type of driving tendency. The standardized score of variables (speed, entry times (for the last-to-go vehicles), headway (for car-following vehicles, headway time is less than 7s [32]) reaction time, the distance from the stop line, and acceleration/deceleration) as input variables for the classification of driving tendency. Results of driving tendency type proportions are 28.35% for conservative, 42.78% for normal, and 28.87% for aggressive ones. The final descriptive statistics are shown in Table 1.
3.2. Driving Variables Analysis

3.2.1. Speeds. Vehicular speed refers to point speed with flashing green onset. The descriptive statistics of different driving tendencies are shown in Table 1, and all of them followed a normal distribution. K-independent sample test was used to determine whether vehicular speed comes from the same distribution of the population. The Kruskal-Wallis test showed 492.59 for the mean rank of the aggressive type, 410.21 for the mean rank of the normal type, and 287.69 for the mean rank of the conservative type, respectively. Test statistics show that asymptotic significance is 0.000, which revealed that speeds of different driving tendencies had significant differences. Additionally, the speeding rates of the aggressive, normal, and conservative type are 7.27%, 2.27%, and 0.45%, respectively, with a speed limit of 60 km/h (16.67 m/s).

3.2.2. Acceleration/Deceleration. Acceleration refers to the average acceleration of the last-to-go vehicle from the flashing green onset to entering the intersection. It can be calculated by the following equation: \( \alpha = 2(X_c - vt)/t^2 \), where \( X_c \) denotes the distance between the last-to-go vehicle to the stop line; \( v \) represents the last-to-go vehicle’s speed (m/s); and \( t \) is the driving time of the last-to-go vehicle.

Deceleration refers to the average deceleration of the first-to-stop vehicle. It can be calculated by the following equation: \( d = v/t \), where \( v \) represents the first-to-go vehicle’s speed (m/s) and \( t \) is the time from brake light onset to the vehicle stopping. The EasyFit software was used to fit the distribution of the acceleration/deceleration to different driving tendencies. The results revealed that the acceleration/deceleration all follow normal distribution but have significant differences (for acceleration, Chi-Square=194.021, P<0.0001; for deceleration, Chi-Square=334.070, P<0.0001).

The distribution probability density is expressed as

\[
f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

where \( \sigma \) is the continuous scale parameter (\( \sigma > 0 \)); \( \mu \) is the continuous location parameter. The calibration of the parameters is as shown in Table 2.

### Headways

Headways were classified into several groups according to different speed ranges (from 0 to 70 km/h, at 10 km/h intervals) and the analysis of K-independent sample tests showed that headways had significant differences under the condition of different driving tendencies (Chi-Square=151.11, P<0.0001). By using EasyFit software, the results of the headways’ distribution under the condition of different driving tendencies revealed that the headways all followed Weibull (3P) distribution. The distribution probability density is expressed as

\[
f(x) = \frac{\alpha}{\beta} \left( \frac{x-\gamma}{\beta} \right)^{\alpha-1} e^{-((x-\gamma)/\beta)^\alpha}
\]

where \( \alpha \) is a continuous shape parameter (\( \alpha > 0 \)); \( \beta \) is a continuous scale parameter (\( \beta > 0 \)); \( \gamma \) is a continuous location parameter (\( \gamma \equiv 0 \) yields the two-parameter Weibull distribution). The calibration of the parameters is as shown in Table 2.

### Probability of Drivers’ Stop/Go Decision

To understand drivers’ decision during the phase transition interval and to determine the boundaries of type II DZ, a binary logistic regression was presented. Before the model coefficients were calibrated, the Pearson correlation coefficient formula was used to test the correlation of the variables to reduce the repetitive effect of the variables on the model. We found that the correlation coefficients between the speeds and headways are greater than 0.6, which belongs to strong correlation. Therefore, the headways variable is deleted. The model can be expressed as follows:

\[
\text{Logit} (P_{Y=1}) = \ln \left( \frac{P_{Y=1}}{1 - P_{Y=1}} \right) = \alpha_0 + \alpha_1 D + \alpha_2 V + \frac{\alpha_3 A}{D} + \alpha_4 E + \alpha_5 S
\]

\[
+ \alpha_6 CF + \alpha_7 T
\]

where \( P_{Y=1} \) denotes the probability that a driver makes a go decision at yellow onset; \( \text{Logit} (P_{Y=1}) \) is the odds ratio of \( P_{Y=1} \); \( D \) is the distance from the stop line (m); \( V \) represents the speed of the vehicle (km/h); \( A/D \) denotes acceleration/deceleration (m/s²); \( E \) denotes the enter time (s); \( S \) denotes the stop time (s); \( CF \) denotes whether a vehicle is in a car-following situation (\( CF=1 \)) or not (\( CF=0 \)); \( T \) is a dummy variable and denotes the driving tendency of a driver, \( T=1 \) represents aggressive type, \( T=2 \) represents normal type, and \( T=3 \) represents conservative type; and \( \alpha_0 \sim \alpha_7 \) are estimated parameters. The results of independent variables significant test show that some explanatory variables, such as entry time, stop time, and car-following situation are not significant. Therefore, these variables were removed from the model.

<table>
<thead>
<tr>
<th>Driving tendencies</th>
<th>Acceleration</th>
<th>Deceleration</th>
<th>Headway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>Normal</td>
<td>Weibull (3P)</td>
</tr>
<tr>
<td></td>
<td>( \sigma )</td>
<td>( \mu )</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>Aggressive</td>
<td>0.565</td>
<td>0.221</td>
<td>0.276</td>
</tr>
<tr>
<td>Normal</td>
<td>0.926</td>
<td>1.698</td>
<td>0.222</td>
</tr>
<tr>
<td>Conservative</td>
<td>0.946</td>
<td>0.052</td>
<td>0.291</td>
</tr>
</tbody>
</table>
The final logistic regression models of drivers’ stop/go decision can be presented by
\[
\ln \left( \frac{P_{Y=1}}{1 - P_{Y=1}} \right) = 4.756 - 0.18D + 0.103V - 0.444T \quad (4)
\]

As shown in (4), the distance from the stop line and the driving tendency negatively affect a driver’s go decision. The farther the distance, the lower the aggressive tendency, as well as the probability of a driver making a stop decision. In contrast, speed and acceleration are positive factors, suggesting that for vehicles with higher speed or higher acceleration the drivers are more prone to make go decisions and cross the intersection.

3.4. Determining the DZ Boundary. According to (4), the logistic regression models of drivers’ stop/go decision can be transformed as
\[
P_{Y=1} = \frac{\exp(4.756 - 0.18D + 0.103V - 0.444T)}{1 + \exp(4.756 - 0.18D + 0.103V - 0.444T)} \quad (5)
\]

Based on type II DZ, we know the upper boundary of the DZ is located where the vehicles have 90 percent probability of stopping and the lower boundary is at 10 percent. In addition, \( T \) is a dummy variable and denotes the driver’s driving tendency: \( T=1 \) represents aggressive type, \( T=2 \) represents normal type, and \( T=3 \) represents conservative type. Therefore, we can put \( P_{Y=1} = 0.9, P_{Y=1} = 0.1, \) and \( T = (T=1 \text{ or } T=2 \text{ or } T=3) \) into (5) and the boundaries of the DZ will be calculated as follows:

① Under the condition of aggressive driving tendency
\[
D_{\text{upper}} = 0.57V + 36.16 \quad (6)
\]
\[
D_{\text{lower}} = 0.57V + 11.75 \quad (7)
\]
② Under the condition of normal driving tendency
\[
D_{\text{upper}} = 0.57V + 33.694 \quad (8)
\]
\[
D_{\text{lower}} = 0.57V + 9.283 \quad (9)
\]
③ Under the condition of conservative driving tendency
\[
D_{\text{upper}} = 0.57V + 29.556 \quad (10)
\]
\[
D_{\text{lower}} = 0.57V + 5.3 \quad (11)
\]

4. Method

As discussed, when the phase transition interval starts, drivers in DZ need to decide to “go” or “stop” at first and then take the appropriate action (such as selecting a perceived safe acceleration/deceleration rate). A rear-end accident is likely to occur when a leading vehicle makes a stop decision and decelerates and the target vehicle fails to stop in the available stopping distance (scene 1: \( P_L \) denotes the probability) or continues driving (scene 1: \( P_L \) denotes the probability).

4.1. Conditional Probability Model

Scene 1. According to Fu [26], rear-end collisions occur during the phase transition interval based on three premises: (1) the leading vehicle is in DZ; (2) the leading vehicle decides to stop and decelerates; and (3) the following vehicle fails to stop in the available stopping distance. Therefore, \( P_1 \) can be estimated using
\[
P_1 = P \left(L_{DZ}\right) \cdot P \left(L_S\right) \cdot P \left(d_i \mid L_S\right) \cdot P \left(F \mid L_S d_i\right)
\]
\[
+ P \left(L_{DZ}\right) \cdot P \left(L_S\right) \cdot P \left(d_i \mid L_S\right) \cdot P \left(F \mid L_S d_i\right)
\]
\[
+ \cdots + P \left(L_{DZ}\right) \cdot P \left(L_S\right) \cdot P \left(d_i \mid L_S\right) \cdot P \left(F \mid L_S d_i\right)
\]
\[
\cdot P \left(F \mid L_S d_i\right) = P_1
\]
\[
= P \left(L_{DZ}\right) \cdot P \left(L_S\right) \cdot \sum_{i=1}^{n} P \left(d_i \mid L_S\right) \cdot P \left(F \mid L_S d_i\right)
\]

\( P \left(L_{DZ}\right) \) denotes the probability of a leading vehicle of being in the DZ; \( P \left(L_S\right) \) denotes the probability that a leading vehicle decides to stop; \( P \left(d_i \mid L_S\right) \) denotes the probability of deceleration \( d_i \) taken by the leading vehicle when it decides to stop; and \( P \left(F \mid L_S d_i\right) \) denotes the probability of the target vehicle failing to avoid collision under the condition that the leading vehicle decides to stop and decelerates.

In (12), we know \( P \left(L_S\right) = 1 - P_{Y=1} \). The estimation of others is as follows.

(i) To Estimate the Probability of the Leading Vehicle of Being in the DZ. When the distance between the leading vehicle and the target vehicle is less than the distance between the following vehicle and the lower boundary of the DZ, it indicates that the leading vehicle is in the DZ.
\[
P \left(L_{DZ}\right) = P \left(\frac{v_f t_h - D_f - D_{\text{lower}}}{v_f}\right)
\]
\[
= P \left(\frac{t_h \leq \frac{D_f - D_{\text{lower}}}{v_f}}{v_f}\right) \quad (13)
\]
where \( v_f \) is the target vehicle’s speed; \( t_h \) denotes the headway between the leading vehicle and the target vehicle; \( D_f \) denotes the distance between the target vehicle and the stop line; and \( D_{\text{lower}} \) is the lower boundary of DZ. \( t_h \) follows the Weibull (3P) distribution (see Table 1). The probability that a leading vehicle is in the DZ under different driving tendency conditions is as follows:

① Under the condition of aggressive driving tendency
\[
P \left(L_{DZ}\right) = \int_{0}^{(D_f - D_{\text{lower}})/v_f} \frac{2.193 \left((x - 0.534)/1.357\right)^{1.193} e^{-((x - 0.534)/1.357)^2}}{1.357} dx \quad (14)
\]
Under the condition of normal driving tendency

\[ P(L_{DZ}) = \int_{0}^{\left| D_{f} - D_{lower} \right| / V_f} \frac{2.584 \left( (x - 0.542) / 1.640 \right)^{1.584}}{1.640} e^{-\left((x-0.542)/1.640\right)^{2.584}} dx \]  

(15)

Under the condition of conservative driving tendency

\[ P(L_{DZ}) = \int_{0}^{\left| D_{f} - D_{lower} \right| / V_f} \frac{2.031 \left( (x - 0.850) / 1.312 \right)^{1.031}}{1.312} e^{-\left((x-0.850)/1.312\right)^{2.031}} dx \]  

(16)

(2) To Estimate the Probability of the Target Vehicle Failing to Avoid a Collision. The relative positions of two adjacent vehicles before and after the leading vehicle brakes and the target vehicle responding accordingly are shown in Figure 3(a). The initial distance between the two vehicles is \( h_0 \), and then the leading vehicle brakes and moves a distance of \( S_1 \) in \( t \) seconds. Meanwhile, the following vehicle driver perceives, reacts, and moves a distance of \( S_2 \). The final distance between the two vehicles becomes \( h \). If \( h > 0 \), it is safe; otherwise, a rear-end accident occurs.

When the leading vehicle brakes, the performance of the driver in the following vehicle consists of three successive components as shown in Figure 3(b): perception of the leading vehicle braking, the brake decision after the perception, and the action of braking. The first two components correspond to the driver’s reaction time \( (t_r) \), set as 0.7s for the aggressive type, 0.9s for the normal type, and 1.1s for the conservative type, respectively. When a driver takes an action, it also takes time for the vehicle to reach its maximum deceleration and finally stop. This includes the brake system response \( (t_b) \), which is set as 0.65s; the time needed to reach the maximum deceleration \( (t_u) \), which is set as 0.2s; and the braking time until the following vehicle stops \( (t_c) \).

In an interval \( t \), the distance of the leading vehicle moves with a deceleration \( d_1 \), which is expressed as

\[ S_1 = v_1 t - \frac{1}{2} d_1 t^2 \]  

(17)

The distance moved by the following vehicle \( S_2 \) is expressed as

\[ S_2 = \begin{cases} 
 v_2 t \\
 v_2 (t_r + t_b) + 0.5 v_2 (t - t_r - t_b) \\
 v_2 (t_r + t_b) + 0.5 v_2 t_u + v_2 (t - t_r - t_b - t_u) - 0.5 d_2 (t - t_r - t_b - t_u)^2 
\end{cases} \]  

(18)

Then, the final distance \( h \) between two adjacent vehicles is expressed as follows:

\[ h = h_0 + S_1 - S_2 - l_1 \]  

(19)

\[ h_0 = v_2 t_b \]  

(20)

The maximum deceleration of \( d_1 \) is determined by the product of the wheel-road adhesive coefficient \( (\mu = 0.8) \), when the road surface is dry) and gravity acceleration \( (g = 9.8m/s^2) \). Thus, \( d_{max} = \mu g = 0.8 \times 9.8 = 7.84m/s^2 \).

Considering the worst situation, the critical deceleration of the leading vehicle is

\[ d_c = \frac{2 v_2 t_h + 2 v_1 (t_c + 1.33) - 2 v_2 (t_c + 1.23) + 7.84 t_c^2 - 2 l_1}{(t_c + 1.33)^2} \]  

(21)

The required minimum deceleration \( (d_r) \) of the leading vehicle to stop safely before the red signal starts as

\[ d_r = \frac{2 (v_1 \tau - D)}{\tau^2} \]  

(22)

where \( \tau \) is remaining time until red starts; \( D \) is the distance from the leading vehicle to the stop line.

Assume \( \xi \) to be the random variable of the leading vehicle decelerating and the deceleration is a continuous variable following a probability density function \( f(x) \). \( P(\xi \mid \psi = x) \), which denotes the probability that the target vehicle fails to avoid a collision under a certain deceleration of the leading vehicle. \( P(\xi \mid \psi = x) = 1 \) if the leading vehicle’s deceleration exceeds \( d_c \); otherwise \( P(\xi \mid \psi = x) = 0 \). Therefore, the conditional probability of failure to avoid a collision is expressed by

\[ \sum_{i=1}^{N} P(d_i \mid L_g) \cdot P(F \mid L_g d_i) \]  

\[ = \int f(x) P(\xi \mid \psi = x) dx = \int_{\max(d_{max},d_r)} f(x) dx \]  

(23)

Given the normal distribution of deceleration discussed previously, the probability of the target vehicle fails to avoid
a collision under different driving tendency conditions as follows:

1. Under the condition of aggressive driving tendency

\[
\sum_{i=1}^{N} P(d_i \mid L_S) \cdot P(F \mid L_S d_i) = \int_{\text{max}(d_i, d_o)}^{d_{\text{max}}} \frac{1}{0.276 \cdot \sqrt{2\pi}} e^{-(x+3.019)^2/(2\times0.276^2)} \, dx
\]  

2. Under the condition of normal driving tendency

\[
\sum_{i=1}^{N} P(d_i \mid L_S) \cdot P(F \mid L_S d_i) = \int_{\text{max}(d_i, d_o)}^{d_{\text{max}}} \frac{1}{0.222 \cdot \sqrt{2\pi}} e^{-(x+2.267)^2/(2\times0.222^2)} \, dx
\]  

3. Under the condition of conservative driving tendency

\[
\sum_{i=1}^{N} P(d_i \mid L_S) \cdot P(F \mid L_S d_i) = \int_{\text{max}(d_i, d_o)}^{d_{\text{max}}} \frac{1}{0.291 \cdot \sqrt{2\pi}} e^{-(x+1.423)^2/(2\times0.291^2)} \, dx
\]

Based on the previous work, the overall model for scene 1 can be written as follows:

\[
P_1 = \left(\frac{1}{1 + e^{\text{Logit}(P_{Y=1})}}\right) \cdot \left(\int_{0}^{(D_f-D_{\text{lower})}/v_f} \frac{\alpha(x-y)}{\beta} \left(\frac{x-y}{\beta}\right)^{\alpha-1} e^{-((x-y)/\beta)^{\alpha}} \, dx\right) \cdot \left(\int_{\text{max}(d_i, d_o)}^{d_{\text{max}}} \frac{1}{\sqrt{2\pi}\sigma} e^{-((x-x_0)/2\sigma)^2} \, dx\right)
\]

Scene 2. The subject vehicle decides to cross the intersection while its leading vehicles decided to stop. Under this condition, the risk probability of a rear-end collision can be estimated using

\[
P_2 = P(L_{DZ}) \cdot P(L_S) \cdot P(F_R) = P(L_{DZ}) \cdot (1 - P_{Y=1}) \cdot P_{Y=1}
\]

\[
= \int_{0}^{(D_f-D_{\text{lower}})/v_f} \frac{\alpha(x-y)}{\beta} \left(\frac{x-y}{\beta}\right)^{\alpha-1} e^{-((x-y)/\beta)^{\alpha}} \, dx
\]

\[
= \left(\frac{e^{\text{Logit}(P_{Y=1})}}{1 + e^{\text{Logit}(P_{Y=1})}}\right) \cdot \left(\frac{1}{1 + e^{\text{Logit}(P_{Y=1})}}\right)
\]

where \(P(F_R)\) denotes the probability of the following vehicle continuing to drive when the leading vehicle makes a stop decision and decelerates: \(P(F_R) = P_{Y=1}\).

The overall risk probability of a rear-end collision is shown in

\[
P = P_1 + P_2
\]

4.2. Monte Carlo Simulation. Calculating the \(P\) value directly is complicated. Therefore, a Monte Carlo simulation approach has been adopted using MATLAB. The Monte Carlo simulation technique produces samples generated from given probability distributions [1]. The average of the output variable (rear-end collision probability in this paper) can be used as an estimator of system evaluation. Twenty thousand pairs of vehicles (leading vehicle and target vehicle) were simulated under the condition of different driving tendencies. Each simulation was randomly assigned a distance, a driving tendency type, a reaction time, an entering time, speed, and headway according to the following rules:
(i) **Distance from the stop line**: The leading vehicle was randomly assigned a distance from 0 to 80m with 0.5m intervals and the target vehicle’s distance was set by adding the result of multiplying its current speed and headway to the leading vehicle’s distance.

(ii) **Speed**: The vehicle’s speed was given according to the speed distribution under different driving tendencies (see Table 1). For the speed of the target vehicles, because of the influence from the leading vehicles, an adjusting factor following normal distribution $N(0.9, 0.27)$ was adopted according to the statistical results of observed data.

(iii) **Headway**: Headways followed Weibull (3P) distribution. The values of the parameters of the distribution are shown in Table 2.

(iv) **Acceleration/Deceleration**: Acceleration/deceleration followed normal distribution. The maximum acceleration and deceleration are $4.8\text{m/s}^2$, $-7.84\text{m/s}^2$, respectively.

5. **Results**

Estimation results of rear-end collision probability under different driving tendencies are shown in Figure 4.
It can be inferred from Figure 4(a) that rear-end collision probability has the lowest value near the stop line at yellow onset and rises as the distance from the stop line increases, which is common for each driving tendency. The difference lies in the following: (1) when aggressive drivers is approaching an intersection at green flashing onset, the rear-end collision potential in a zone of 20m–40m is well-matched with that of conservative drivers. However, if it is in the zone above 40m, the collision potential is lower than that of the conservative type; (2) the rear-end collision probability of the conservative type of driver increases sharply in a zone of 40m–70m; and (3) the rear-end collision probability of normal type of drivers always increases steadily, and it is always the lowest compared with the other two types. The rear-end collision probability for scene 2 is shown in Figure 4(b). In the cases, the rear-end collision probability for the aggressive type is the highest, followed by that of the conservative and normal driver type.

The total collision rear-end probability of different driving tendency is shown in Figure 4(c). We can see that the collision rear-end probability is different by different driving tendency even at the same location. Aggressive drivers, because of the shorter headway, higher speed, and aggressive acceleration or abrupt stops, have the highest rear-end collision probability during the transition interval; and conservative drivers, because of their lengthy reaction time and abrupt stop action from being overcautious, are no safer than normal drivers. In addition, we found that the DZ shifts towards the stop line (compare the zone where increases sharply) when the driving tendency changes from the aggressive type to the conservative type to the normal type. This indicates that the rear-end collision probability under each driving tendency does not totally correspond to the DZ distribution (see (6)–(11)).

6. Conclusions

The DZ problem is a leading cause for rear-end accidents at signalized intersections. In this paper, based on field observations and over 300 hours of recorded video data collection at two intersections and data analysis, we found that the collision rear-end probability is different even if a vehicle is at the same position with the same speed because the driver’s tendency is a key factor that cannot be ignored. To discover how the driving tendency impacts the DZ distribution and then impacts the probability of rear-end accidents, we proposed a conditional probability model of rear-end accidents and developed a Monte Carlo simulation framework to calculate the model. The conclusion that the aggressive type is the highest collision rear-end probability followed by the conservative and normal is drawn. However, to draw general conclusions, observations and data collection are needed for different types of intersections with various geometric settings, traffic compositions, and signal timings.

Data Availability

The trajectory data used to support the findings of this study have not been made available because the National Science Foundation of China (No. 51478110) has not completed them.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 51478110, 51508122) and Science and Technology Program of Jiangsu Province (No. BY2016076-05).

References


