Survival analysis of lane-changing duration characteristic for heavy vehicles and passenger cars using the HighD dataset

Yang Li, Linbo Li*, Daiheng Ni

Yang Li
Ph.D., Candidate
Key Laboratory of Road and Traffic Engineering of Ministry of Education,
Tongji University, China,
4800 Cao'an Road, Shanghai, 201804,
E-mail: cc960719@tongji.edu.cn

Linbo Li
Ph.D., Associate Professor
Key Laboratory of Road and Traffic Engineering of Ministry of Education,
Tongji University, China,
4800 Cao'an Road, Shanghai, 201804
E-mail: llinbo@tongji.edu.cn

Daiheng Ni
Ph.D., Professor
Civil and Environmental Engineering,
University of Massachusetts Amherst, Massachusetts 01003, USA
E-mail: ni@engin.umass.edu

Abstract

The LC (lane-changing) behavior of heavy vehicles has a more profound impact on traffic flow in contrast to passenger cars. This paper investigates LC (lane-changing) duration characteristic for heavy vehicles and passenger cars using the HighD dataset. LC duration is an indispensable indicator to characterize LC behavior, which measures the total time it takes for a vehicle to complete the LC maneuver. LC trajectories are extracted from the HighD dataset, which contains of 16.5 hours of measurement and over 11,000 vehicles. The univariate and regression survival analysis approach are introduced to explore the overall survival function and the influencing factors. The results suggest that a significant difference exists in LC behavior characteristic between these two types of vehicles, with the survival curve of heavy vehicles always above that of passenger cars. These two curves show a trend of moving away from each other and then gradually approaching. The MST (Median Survival Time) of heavy vehicles is 0.57s higher than passenger cars. Also, we find that the heavy vehicles are more inclined to perform LC when their speed is close to a certain speed value (possibly desired or maximum speed). Their LCD are less susceptible to the time-headway and distance-headway and more susceptible to their own speed. Finally, the main findings, modeling implications, practical applications, and future work have been discussed. We hope this paper could contribute to our further understanding of LC behaviors of heavy vehicles and passenger cars.

Keywords: Lane-changing behavior, Lane-changing duration, Heavy vehicles, Survival analysis, HighD dataset.

1 Introduction

As one of the important components of road traffic, heavy vehicles run on thousands of kilometers of highways, providing material protection for the normal operation of national economies. Although heavy vehicles make up only a small portion of the traffic flow, they exert a far more significant impact on traffic flow than passenger cars. This is mainly due to their physical and psychological effects on surrounding vehicles (Moridpour et al., 2015). These effects are the results of their larger physical state (length, width, mass), and their significant differences in operating characteristics (acceleration, deceleration, maneuverability) with passenger cars (Assemi and Hickman, 2018; Cao et al., 2016; Castillo-Manzano et al., 2021; Moridpour et al., 2015). Studies have shown that the severity of traffic accidents caused by heavy vehicles is much higher than that of passenger cars, and the rising motorization rates for trucks is more likely to cause higher traffic fatalities (Assemi and Hickman, 2018; Castillo-Manzano et al., 2015). Therefore, many scholars have gradually begun to focus on the analysis of the driving behavior of heavy vehicles in recent years, so as to grasp traffic flow characteristic of the m more comprehensively, as well as to help guide the modeling of heavy vehicles in traffic simulation software.

The focus of existing research has tended to be on two basic driving behaviors that could be frequently observed on the road, one of which is the CF (car-following) behavior and the other is the LC (lane-changing) behavior. CF behavior describes the operation of the vehicle in the longitudinal direction during forward travel. LC behavior describes the lateral movement of the vehicle from the current-lane to the target-lane while proceeding forward. Since more interaction with surrounding vehicles are simultaneously involved, the impact of LC maneuver on traffic flow is more pronounced than that of CF maneuver (Moridpour et al., 2015). Extensive studies have shown significant differences in driving behavior between heavy vehicles and passenger cars (Aghabayk et al., 2011; Aghabayk et al., 2014a; Cao et al., 2016; Chen et al., 2016; Ossen and Hoogendoorn, 2011; Yang et al., 2019). Ossen and Hoogendoorn (2011) found that compared with the passenger cars, the heavy vehicles drivers are more likely to adopt a more "robust" CF behavior. Chen et al. (2016) found that heavy vehicles tend to reduce speed changes triggered by traffic disturbances, while passenger cars following heavy vehicles tend to amplify these disturbances. Aghabayk et al. (2014b) found that the variables affecting the driver's acceleration in the CF behavior also vary depending on the CF combination. Moridpour et al. (2010) investigated the effect of surrounding traffic characteristics on the LC behavior between the passenger cars and heavy vehicles. Results suggested a substantial difference of LC behaviors exists between the passenger cars and heavy vehicles. Aghabayk et al. (2011) compared the LC maneuvers between the heavy vehicles and passenger cars on arterial roads and freeways. Results indicated that the type and the size of vehicles influence the LC maneuver, in particular on arterial road. Cao et al. (2016) analyzed the difference of LC behaviors between heavy vehicles and passenger cars. A model framework for the execution of LC with regard to the emergency status and the impact of surrounding traffic for individual drivers is established.

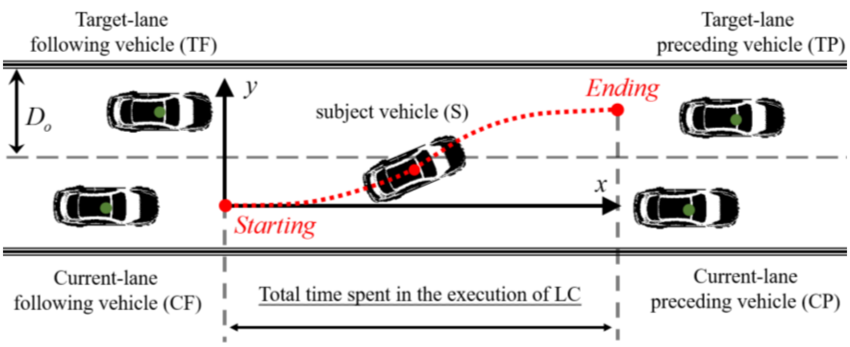


Figure 1 The schematic diagram of LC behavior (dot curve represents the LC trajectory from the starting point to the ending point)

Despite the fruitful results of existing studies, less attention was paid on the investigation of the LC duration, especially for heavy vehicles. LCD (LC duration) measures the total time spent in the execution of LC as shown in Figure 1, which is one of the most important characteristics for us to comprehend the LC behavior. Toledo and Zohar (2007) employed the multiple linear regression model to analyze the LCD of the NGSIM dataset. Results demonstrated that traffic density, by the direction of the change, and by other vehicles around the subject vehicle may affect the LCD. Wu et al. (2008) adopted the semi-parametric survival analysis model to analyze the mandatory LCD data. Results demonstrated that there is no significant evidence showing that different vehicle types have an effect on LCD, but there is a significant difference in LCD during different time periods. Vlahogianni (2013) utilized three AFT (Accelerated Failure Time) models to analyze the overtaking duration in two-lane highways. Results demonstrate the speed difference relative to the preceding vehicle, the speed of opposing traffic, the spacing from the lead and opposing traffic, and the driver's gender may influence the LCD. Wang et al. (2014) found that there is no significance difference between the left-to-right LCD. They also conjectured that the duration times will reach a saturation value, when the velocity becomes even higher. Li et al. (2021) presents a comprehensive analysis of LCD from the perspective of survival analysis. LC trajectories are extracted from the HighD dataset. Both comparative univariate and regression survival analysis of LCD have been carried out.

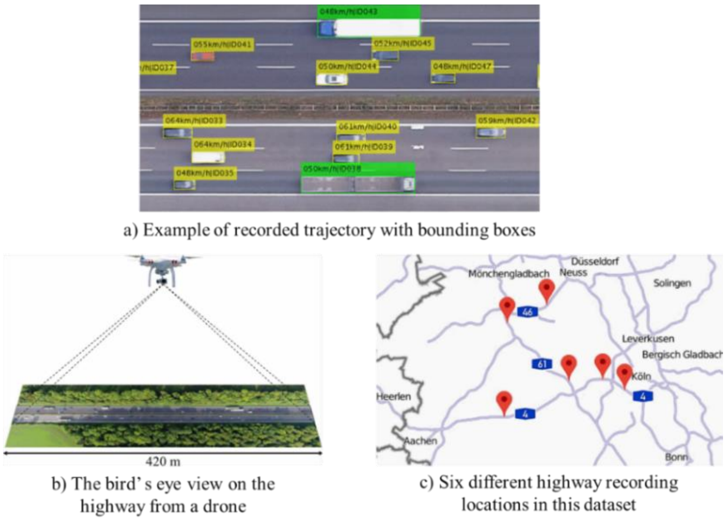
Although above-mentioned studies have yielded some results, the majority of them concentrate on passenger cars, while overlooking the heavy vehicles. However, heavy vehicles' LC behavior is significantly different from passenger cars, and has more obvious impact on the real traffic flow (Cao et al., 2016; Toledo and Zohar, 2007). Such negligence may inevitably lead us to a one-sided understanding of LC behavior of microscopic traffic flow characteristic. On the basis of our previous research (Li et al., 2021), this paper aims to take a further step to explore the LCD characteristic for heavy vehicles and passenger cars. Meanwhile, the HighD dataset is adopted in this paper (Krajewski et al., 2018), which is a new dataset of naturalistic vehicle trajectories. In contrast to the datasets used in the existing literature, such as the NGSIM dataset (Aghabayk et al., 2011; Toledo and Zohar, 2007; Wang et al., 2014), this dataset has a higher acquisition accuracy and larger data size (Krajewski et al., 2018). More importantly, the percentage of heavy vehicles in this dataset tops out at 23%, much higher than the 3% in the NGSIM dataset (Krajewski et al., 2018). Therefore, this dataset is more suitable for in-depth analysis of LCD for heavy vehicles. These are the main contributions of this paper. The obtained findings and modeling implications may help us have a more comprehensive understanding of LC behaviors. The remainder of this paper is organized as follows. Section 2 presents the description and processing of the HighD dataset, and some preliminary analysis. Section 3 and Section 4 present the in-depth analysis of LCD through the survival analysis approach. Finally, the conclusion is presented in Section 5.

2 Data description and processing procedures

The HighD dataset is employed in this paper is recorded on German highways during 2017 and 2018 (Krajewski et al., 2018). This dataset contains of 16.5 hours of measurement, 45,000 kilometers of total driven distance and over 11,000 vehicles. These trajectories are recorded in 4k (4096*2160) resolution from six different locations near Cologne, Germany. At the same time, the positioning error of each trajectory is typically less than ten centimeters. For more details of this paper, please refer to Krajewski et al. (2018). Figure 2 presents the brief introduction of the HighD dataset, including the bounding boxes of each vehicle, the bird’s eye view on the highway, and six different highway recording locations. Table 1 presents the data format of each track file in the HighD dataset, including the id, position, speed, acceleration, and the id of surrounding vehicles for each vehicle. The interval between each two records is about 0.025s.

The process of data processing and trajectory extraction is as follows. First, we merge a total of 60 tracks files. Second, we exclude the data with failed LC and missing trajectories. Finally, we determine the beginning and ending points of LC for each vehicle. Similar to our recent study (Li et al., 2021), these two points are identified with reference to three main variables: lateral position, lateral velocity, and lateral acceleration. An example of discriminating these two points of LC trajectory (The vehicle with id 1384 in the 25th tracks file) is given in Figure 3. The topmost subplot represents the lateral trajectory of the vehicle, which gradually changes from 18m to 22.1m, from which it can be seen that the lane width is roughly 4m. The middle subplot represents the acceleration of the vehicle, which oscillates around 0 when it keeps CF. When the vehicle starts to perform LC, its acceleration gradually increases from 0 to 0.5 m/s^2, then slowly decreases to -0.34 m/s^2, and then gradually increases again to 0 (similar to a sin/cos curve). The lower subplot represents the lateral speed, whose value gradually increases from 0 to a maximum value of 1.2m/s, and then declines gradually to 0m/s. Thus, we choose the moment when the lateral velocity and acceleration of the vehicle equal to 0 as the beginning and ending point of LC, for example, the red dot in Figure 3. It is worth noting that we cannot rely solely on the variable speed, but also need to combine acceleration and position information. This is because the speed and acceleration in the raw dataset are always vibrating around the value of zero.

Finally, a total of 746 heavy vehicles trajectories and 7674 passenger cars trajectories are extracted. Figure 4 present the descriptive statistics of the LC trajectories for heavy vehicles and passenger cars. It could be observed that there is a significant difference between the LC behavior of heavy vehicles and passenger cars. The mean LCD and median LCD of passenger cars is about 5.7s and 5.55s, while the mean LCD and median LCD of heavy vehicles is about 6.22s and 6.08s. The LCD of heavy vehicles is slightly higher than of passenger cars (about 0.5s). The median value is lower than the mean value indicates that half of the vehicles pull up the mean LCD of all the vehicles.



a) Example of recorded trajectory with bounding boxes

b) The bird's eye view on the highway from a drone

c) Six different highway recording locations in this dataset

Figure 2 The brief introduction of the HighD dataset including the bounding boxes of each vehicle, the collecting method and the recording locations

Table 1 Data format for each track file in the HighD dataset

Name	Description
frame	The current frame.
id	The Track's id.
x	The x position of the upper left corner of the vehicle's bounding box.
y	The y position of the upper left corner of the vehicle's bounding box.
width	The width of the bounding box of the vehicle.
height	The height of the bounding box of the vehicle.
xVelocity	The longitudinal velocity in the image coordinate system.
yVelocity	The lateral velocity in the image coordinate system.
xAcceleration	The longitudinal acceleration in the image coordinate system.
yAcceleration	The lateral acceleration in the image coordinate system.
frontSightDistance	The distance to the end of the recorded highway section in driving direction from the vehicle's center.
backSightDistance	The distance to the end of the recorded highway section in opposite direction from the vehicle's center.
dhw	The distance headway.

thw	The time headway.
ttc	The time-to-collision.
precedingXVelocity	The longitudinal velocity of the preceding in the image coordinate system.
precedingId	The id of the preceding vehicle in the same lane.
followingId	The id of the following vehicle in the same lane.
leftPrecedingId	The id of the preceding lane on the left in the direction of travel.
leftAlongsideId	The id of the adjacent lane on the left in the direction of travel.
leftFollowingId	The id of the following lane on the left in the direction of travel.
RightPrecedingId	The id of the preceding lane on the right in the direction of travel.
RightAlongsideId	The id of the adjacent lane on the right in the direction of travel.
RightFollowingId	The id of the following lane on the right in the direction of travel.
laneId	The IDs start at 1 and are assigned in ascending order.

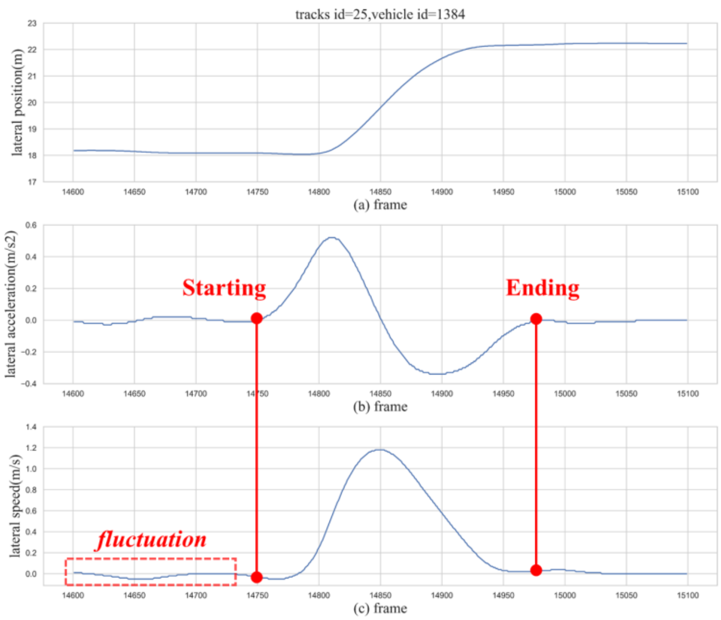


Figure 3 Example of determining the starting and ending point of the LC trajectory

At the same time, the time-headway and distance-headway of heavy vehicles are all higher than that of passenger cars, while the speed of heavy vehicles is much lower than that of passenger cars. The mean time-headway of heavy vehicles and passenger cars is about 2.76s and 1.53s. The mean distance-headway of heavy vehicles and passenger cars is about 69.67m and 45.29m. This indicates that heavy vehicles prefer to maintain greater spacing from the vehicle in front when perform LC. In addition, heavy vehicles have a more concentrated range of speed variation (less variance) than passenger cars. The 25%th percent, 50th percent, and 75th percent of the speed of the heavy trucks are relatively close (around 24.14m/s, 25.03m/s, 25.88m/s respectively). And these three values of passenger cars are 26.45m/s, 30.26m/s, 33/11m/s respectively. This may reveal an important characteristic of LC behavior of heavy vehicles. Since the heavy vehicle has poor acceleration and deceleration characteristics, it prefers to change lanes at a stable speed (probably close to the desired speed) to ensure the safe completion of LC.

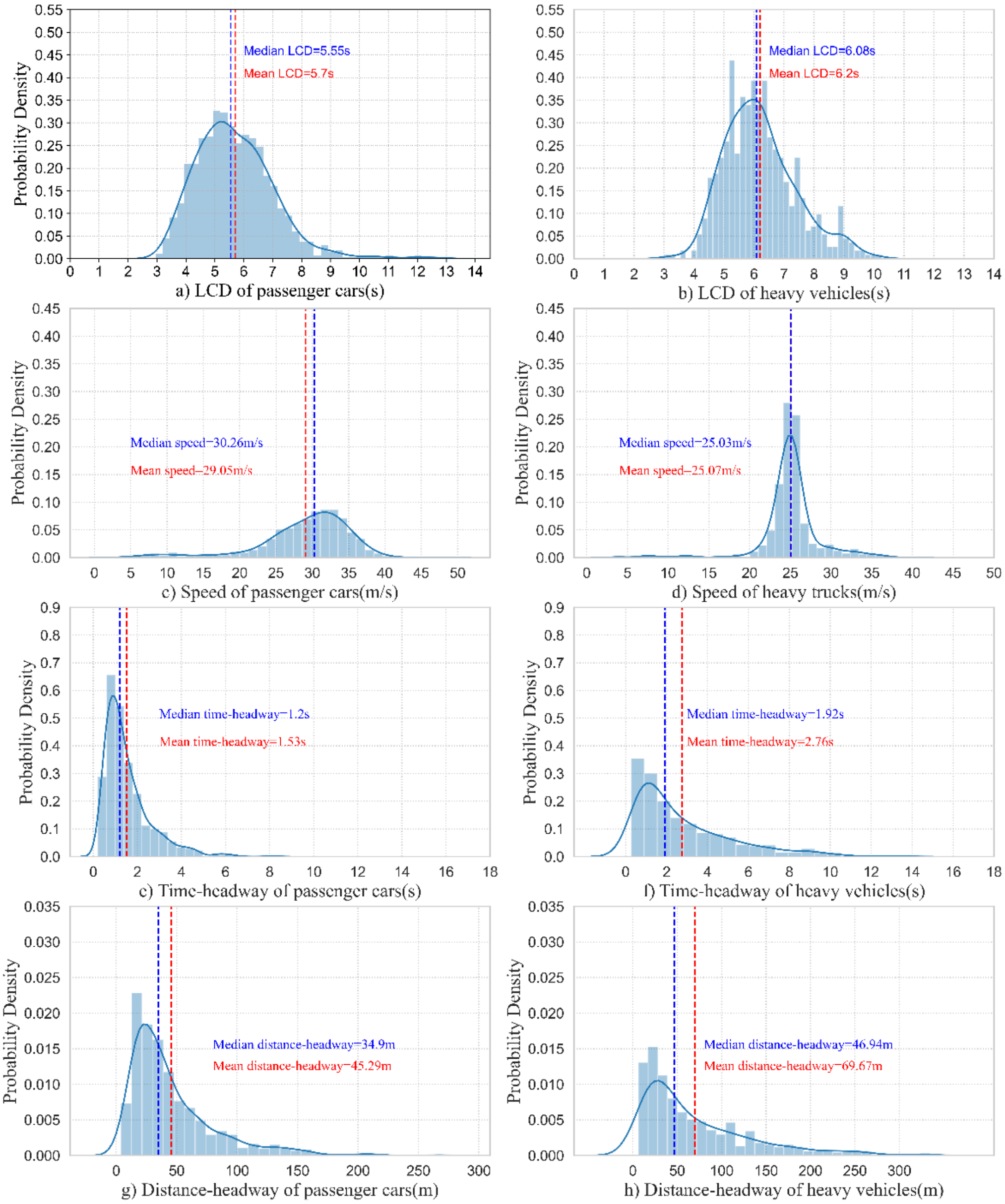


Figure 4 Time-headway and distance-headway distribution of heavy vehicles and passenger cars

3 Survival analysis approach

In general, the survival analysis is a collection of statistical procedures of data analysis for which the outcome variable of interest is time until an event occurs (Kleinbaum and Klein, 2012). And the main goal of survival analysis is to estimate the survival and hazard function of the duration data, and to assess the relationship of explanatory variables to duration time. On the basis of our recent research (Li et al., 2021), we also employ this method to investigate the difference of LCD characteristic for heavy vehicles and passenger cars. The univariant survival model is introduced to explore the overall survival function difference for heavy vehicles and passenger cars. The AFT (Accelerated Failure Time) model is adopted to investigate the influencing factors.

3.1 Univariant survival model

Same as our previous research, five commonly-used survival distribution functions are selected (Li et al., 2021). For basic concepts of survival analysis and detailed formula derivation, please refer to (Kleinbaum and Klein, 2012; Li et al., 2021; Washington et al., 2020). Let T denotes the continuous non-negative random variable representing survival time, and we could transform T into the following form:

$$Y=\log\mathrm{T}=\alpha+\sigma\mathrm{W} \tag{1}$$

The exponential distribution has the constant hazard value, which means the conditional probability of an event is constant over time.

$$h(t)=\lambda \tag{2}$$

Where λ is the constant hazard. Then, we could derive $S(t)=\exp\left(-\lambda t\right)$, $f(t)=\lambda\exp\left(-\lambda t\right)$.

The Weibull distribution assumes the hazard obeys the Weibull distribution, which is capable of allowing for positive, negative, or even no duration dependence.

$$h(t) = p\lambda^p t^{p-1} \quad (3)$$

Where p and λ are the parameters that controls the shape of $h(t)$. The hazard is rising if $p > 1$, constant if $p = 1$, and declining if $p < 1$.

The Lognormal distribution assumes the W has a standard normal distribution.

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{1}{2} (\frac{\log t - \mu}{\sigma})^2 \right) \quad (4)$$

The $h(t)$ increases from 0 to reach a maximum and then decrease monotonically, approaching 0 as $t \to \infty$.

The Loglogistic distribution assumes the W has a standard logistic distribution. The hazard function is given below.

$$h(t) = \frac{\lambda p(\lambda t)^{p-1}}{1 + (\lambda t)^p} \quad (5)$$

Where the hazard itself monotone decreasing from ∞ if $p < 1$, monotone decreasing from λ if $p = 1$, and similar to lognormal if $p > 1$.

The Generalized-Gamma distribution assumes the W has generalized extreme value distribution with parameter k .The density of the Generalized Gamma distribution is formulated as:

$$f(t) = \frac{\lambda p(\lambda t)^{pk-1} e^{-(\lambda t)^p}}{\Gamma(k)} \quad (6)$$

Where $p = 1/\sigma$. The Generalized Gamma includes the following interesting special cases: gamma when $p = 1$, Weibull when $k = 1$, Exponential when $p = k = 1$, Lognormal when $k \to \infty$.

3.2 Accelerated Failure Time model

The AFT model directly models the survival time of the duration data, which assumes the linear relationship between the survival time and the covariates. Three AFT models (LoglogisticAFT, LognormalAFT, WeibullAFT) (Vlahogianni, 2013) are employed to explore the influencing factors of LCD for heavy vehicles and passenger cars. The AFT model can be expressed as:

$$Y_i = x_i^T \beta + W_i \tag{7}$$

Where $Y = \log T$, W_i represents the independent residuals, which is an error term with different density function. Then, we could rewrite the above Equation as:

$$T_i = T_0 \exp(x_i^T \beta) \tag{8}$$

Where $T_0 = \exp(W_i)$. When the $j-th$ dimension changes Δ_j, the survival time would change $\exp(\Delta_j \beta_j)$.

The distribution assumption of the W_i determine which sort of AFT model describes the distribution of the survival time T . The corresponding error term distributions of W_i of the above three models are extreme value, normal and logistic distribution.

4 Results analysis and discussion

4.1 Overall survival function characteristic

Subgraphs (a) and (b) in Figure 5 presents the estimation of the overall survival function of LCD for heavy vehicles and passenger cars under five commonly-used survival distribution functions. Table 2 presents the corresponding AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and MST (Median Survival Time). The AIC and BIC are both metrics of assessing model fit penalized for the number of estimated parameters. BIC penalizes model more for free parameters, and the AIC prefers a more complex over a simpler model. AIC has the danger of over fitting and BIC has the danger of under fitting. Therefore, both AIC and BIC are recommended when choosing the best parameters. It could be found that the

Exponential distribution exhibits the worst performance, while the Generalized Gamma distribution outperforms than other distributions both in AIC and BIC. Therefore, we employ the Generalized Gamma distribution to conduct the subsequent analysis.

Subgraph (c) in Figure 5 presents the comparison of the survival function between heavy vehicles and passenger cars using the Generalized Gamma distribution. The survival function decreased rapidly in 3s~8s, while decreased gently in 8s~12s. With the help of this method, we can also intuitively perceive the significant difference between the LCD of heavy vehicles and passenger cars. The survival probability of heavy vehicles at each time is higher than that of passenger cars. With the increase of timeline, these two curves show a tendency to move away from each other first, and then to approach each other again. MST is defined as the time where on average 50% of the duration has expired, which indicates that each vehicle has a 50% chance of completing its LC maneuver. The MST of heavy vehicles is 0.57s higher than that of passenger cars. Subgraph (d) in Figure 5 presents the at-risk percentage difference for heavy vehicles and passenger cars. The distance between these two curves is the furthest when timeline equals to 5s. The at-risk percentage difference value reaches the highest value, roughly around 17%.

Table 2 Fitting results of five commonly-used distribution form

Parametric estimator	AIC		BIC		MST(s)	
	Vehicle	Truck	Vehicle	Truck	Vehicle	Truck
Weibull	3772.47	2157.11	26244.56	13932.99	5.697	6.26
Exponential	5795.27	3649.04	40341.62	23395.64	3.95	4.31
Lognormal	3514.86	2044.73	24450.76	13206.02	5.54	6.1
Loglogistic	3529.63	2057.15	24553.63	13286.34	5.54	6.07
Generalized Gamma	3511.84	2045.72	24417.83	13201.43	5.51	6.08

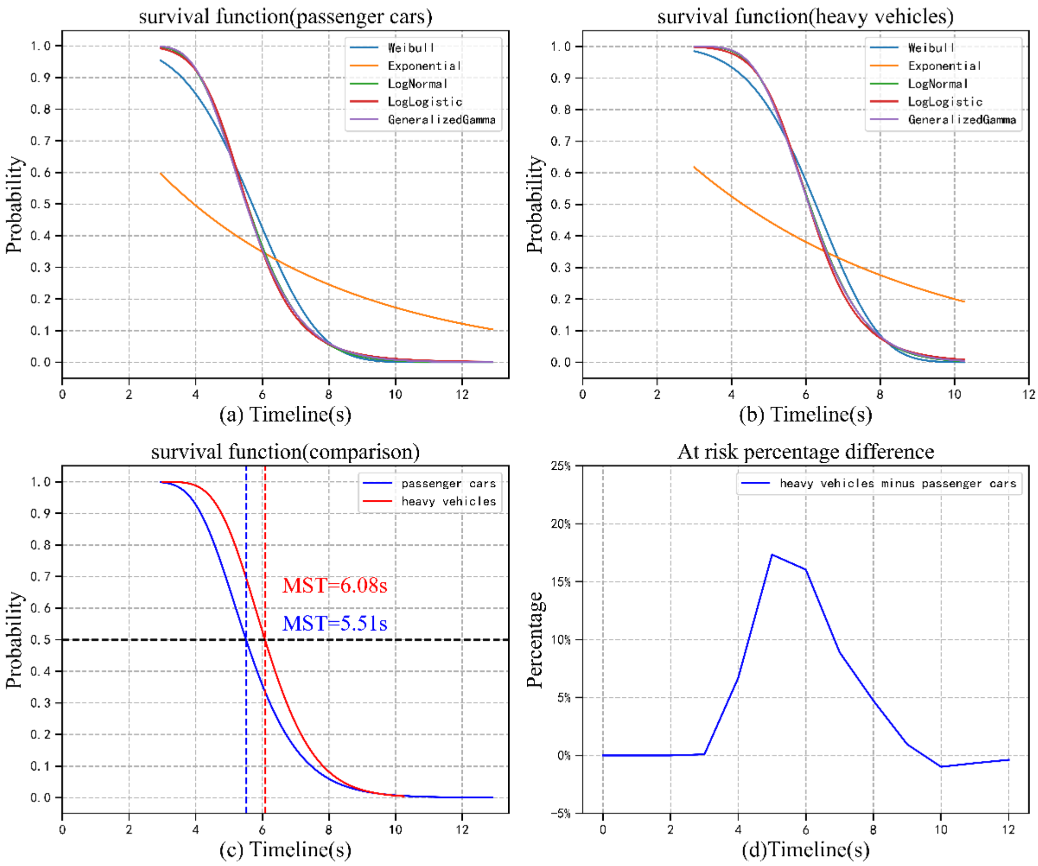


Figure 5 The survival and cumulative hazard function of LCD using the parametric method (five commonly used distribution)

4.2 Influencing factors of LCD

Table 3 and Table 4 presents the regression results using the AFT models. It could be found that these three models achieved relatively close regression results. Compared with the LognormalAFT and the WeibullAFT model, the LoglogisticAFT model has the lowest AIC value. The corresponding MST value of passenger cars and heavy vehicles are 5.48s and 6.04s. Therefore, we employ the LoglogisticAFT to conduct the following analysis. Figure 6 presents the comparison between the baseline survival curve versus the survival curve when some covariates are varied over values. The baseline survival is calculated according to the original dataset. This is useful for us to understand subject's survival as we vary specific covariate. From Figure 6, it could be found that with the increase of the speed, both passenger cars and heavy vehicles will have shorter LCD. With the increase of time-headway, passenger cars are more likely to have a longer LCD, which is opposite to the change direction of distance-headway.

From the regression results, it could be found that all these three variables would significantly affect the LCD of passenger cars (all the p-values are under 0.01). Take the time-headway coefficient of passenger cars for example, a unit increase of the time-headway would result in 17.4% ($e^{0.16}$-$1 = 1.174$-1) increase of baseline survival time T_0. For each additional unit increase of these three variables, sorted by the degree of impact, the baseline survival time of passenger cars would correspondingly increase or decrease by -0.6%, -0.4% and 17.4%. As for heavy vehicles, the p-values of distance-headway and time-headway are all higher than 0.5, while only the p-value of speed is under 0.005. a unit increase of speed would result in the decrease of the baseline survival time by 1.6%.

Table 3 Regression results of LCD of passenger cars

	LoglogisticAFT (AIC=3356.91, MST=5.48s)			LognormalAFT (AIC=3389.19, MST=5.46s)			WeibullAFT (AIC=3419.81, MST=5.64s)		
	coef	exp(coef)	p	coef	exp(coef)	p	coef	exp(coef)	p
Speed	-0.006	0.994	**0.010**	-0.007	0.993	0.005	-0.008	0.992	0.005
Distance-headway	-0.004	0.996	**0.005**	-0.004	0.996	0.005	-0.006	0.994	0.005
Time-headway	0.160	1.174	**0.005**	0.140	1.150	0.005	0.210	1.234	0.005

Table 4 Regression results of LCD of heavy vehicles

	LoglogisticAFT (AIC=1905.16, MST=6.04s)			LognormalAFT (AIC=1911.72, MST=6.05s)			WeibullAFT (AIC=2005.59, MST=6.22s)		
	coef	exp(coef)	p	coef	exp(coef)	p	coef	exp(coef)	p
Speed	-0.016	0.984	0.005	-0.013	0.987	0.005	-0.014	0.986	0.005
Distance-headway	0.001	1.001	0.560	0.000	1.000	0.560	0.000	1.000	0.750
Time-headway	-0.010	0.990	0.780	0.000	1.000	0.920	0.000	1.000	0.980

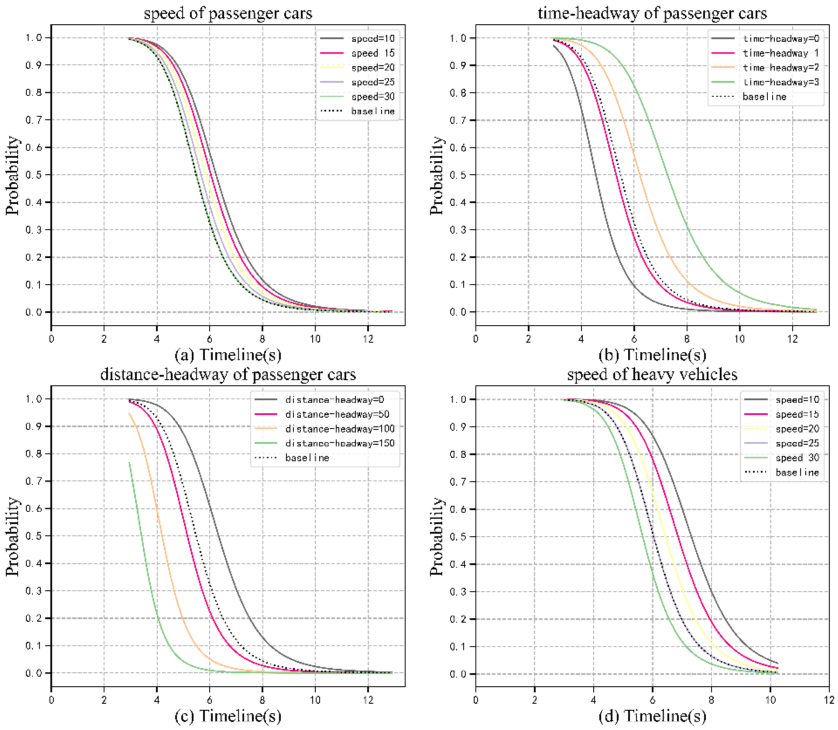


Figure 6 Partial effects of coefficients on the LCD

4.3 Discussion

The center of this paper is to compare and analyze the LCD characteristic between heavy vehicles and passenger cars. On the basis of our recent study (Li et al., 2021), we also employ the HighD dataset and the survival analysis approach to conduct this research. The determination of the starting and ending points of LC trajectory relies on the lateral position, speed and acceleration trajectory of the vehicle. A total of 746 LC trajectories of heavy vehicles and 7674 LC trajectories of passenger cars are extracted from the HighD dataset. After some preliminary analysis of LCD, the univariant model and regression survival model are employed to conduct in-depth analysis. The main findings and implications are summarized below.

We found that there exists a remarkable difference in LC behavior between the heavy vehicles and passenger cars. The average LCD of heavy vehicles and passenger cars of the HighD dataset is about 6.22s and 5.70s. The survival curve of heavy vehicles at each timeline is always above that of passenger cars as shown in Figure 5. These two curves show a trend of moving away from each other and then gradually approaching, with the farthest apart moment being the 5th second. This indicates that the difference between the proportion of passenger cars completing LC and the proportion of heavy vehicles completing LC reaches a maximum at the 5 second (the slope of the survival curve at each moment of passenger cars is greater than that of heavy vehicles). After the 5 second, the survival curve of the heavy vehicles is steeper, and the difference between the ratio of these two completing LC gradually goes to zero. In addition, the 25% percent, median, and 75% percent of the speed of heavy vehicles are all round 25m/s. This reveals that the heavy vehicles are more inclined to perform LC when their speed is close to a certain speed value, which may be the desired or the maximum speed limit. This is in line with the actual driving situation. Due to their physical characteristics (large size and weight, poor acceleration/deceleration characteristics), their low-speed LC behavior is more likely to cause traffic accidents and disrupt traffic flow, so they would choose a higher speed. Meanwhile, we also found that while heavy vehicles would maintain a longer time-headway and distance-headway with preceding vehicle when changing lane, their LCD are less susceptible to such interactions with preceding vehicle and more susceptible to their own speed.

These novel findings may provide some guidance for future traffic modeling. The results obtained based on the HighD dataset could be directly applied to microscopic traffic simulation software to reproduce the execution process of LC behavior of heavy vehicles and passenger cars more realistically. Several metrics could be adopted to assist us explore the differences in LC behavior between different regions or different time periods or different type of drivers. Undoubtedly, many aspects of this paper need further research. Due to the limitations of the dataset, it only records the trajectory information of each vehicle, and lacks the variables such as weather, roadway geometry, driver characteristic, etc. Therefore, one of the important research directions of this paper it to collect LC trajectory data containing the above variables. This may help us further explore the LC characteristic differences between the heavy vehicles and passenger cars. Meanwhile, future research may try to cover a wider range of survival models, and investigate the influencing factors of unsuccessful LC events.

5 Conclusion

The impact of heavy vehicles' LC behavior on traffic flow is significantly higher than that of passenger cars. Nevertheless, existing studies have often overlooked such kind of comparison analysis, especially for LCD characteristics. Therefore, this paper takes a further refinement of the study of the difference in LCD characteristic between heavy vehicles and passenger cars. As one of the indicators to evaluate the execution phase of LC, LCD is an important reference standard for our intuitive understanding to grasp the LC behavior, as well as an indispensable parameter in microscopic traffic simulation software. The main findings, implications, practical applications, limitations, and future work are summarized above. We hope that these findings will help improve our further understanding of LC behavior for heavy vehicles and passenger cars.

6 CRediT authorship contribution statement

Yang Li: Conceptualization, Data curation, Writing - original draft. Linbo Li: Methodology, Funding acquisition, Writing - original draft. Daiheng Ni: Investigation, Writing - review & editing, but with no involvement in the research grant.

7 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8 Data Availability Statement

The HighD dataset is publicly available online. Some of the models, or code generated or used during this study were provided by the National Key R&D Program. They are proprietary or confidential in nature and may only be provided with restrictions.

9 Acknowledgements

This research was funded by the National Key R&D Program of China [grant numbers 2018YFE0102800].

10 Reference

Aghabayk, K., Moridpour, S., Young, W., Sarvi, M., Wang, Y., 2011. Comparing Heavy Vehicle and Passenger Car Lane-Changing Maneuvers on Arterial Roads and Freeways. *Transportation Research Record* 2260(2260), 94-101.

Aghabayk, K., Sarvi, M., Forouzideh, N., Young, W., 2014a. Modelling heavy vehicle car-following behaviour in congested traffic conditions. *Journal of Advanced Transportation* 48(8), 1017-1029.

Aghabayk, K., Sarvi, M., Young, W., 2014b. Attribute selection for modelling driver's car-following behaviour in heterogeneous congested traffic conditions. *Transportmetrica A: Transport Science* 10(5), 457-468.

Assemi, B., Hickman, M., 2018. Relationship between heavy vehicle periodic inspections, crash contributing factors and crash severity. *Transportation Research Part A: Policy and Practice* 113, 441-459.

Cao, X., Young, W., Sarvi, M., Kim, I., 2016. Study of Mandatory Lane Change Execution Behavior Model for Heavy Vehicles and Passenger Cars. *Transportation Research Record* 2561(2561), 73-80.

Castillo-Manzano, J.I., Castro-Nuño, M., Fageda, X., 2015. Can cars and trucks coexist peacefully on highways? Analyzing the effectiveness of road safety policies in Europe. *Accident Analysis & Prevention* 77, 120-126.

Castillo-Manzano, J.I., Castro-Nuño, M., Fageda, X., 2021. Analyzing the safety impact of longer and heavier vehicles circulating in the European market. *Journal of Safety Research* 77, 1-12.

Chen, D., Ahn, S., Bang, S., Noyce, D., 2016. Car-Following and Lane-Changing Behavior Involving Heavy Vehicles. *Transportation Research Record* 2561(1), 89-97.

Kleinbaum, D.G., Klein, M., 2012. *Survival Analysis: A Self-Learning Text, Third Edition*. Springer, New York.

Krajewski, R., Bock, J., Kloeker, L., Eckstein, L., 2018. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems, *2018 21st International Conference on Intelligent Transportation Systems*, pp. 2118-2125.

Li, Y., Li, L., Ni, D., Zhang, Y., 2021. Comprehensive survival analysis of lane-changing duration. *Measurement* 182, 109707.

Moridpour, S., Mazloumi, E., Mesbah, M., 2015. Impact of heavy vehicles on surrounding traffic characteristics. *Journal of Advanced Transportation* 49(4), 535-552.

Moridpour, S., Rose, G., Sarvi, M., 2010. Effect of Surrounding Traffic Characteristics on Lane Changing Behavior. *Journal of Transportation Engineering* 136(11), 973-985.

Ossen, S., Hoogendoorn, S.P., 2011. Heterogeneity in car-following behavior: Theory and empirics. *Transportation Research Part C-Emerging Technologies* 19(2), 182-195.

Toledo, T., Zohar, D., 2007. Modeling Duration Of Lane Changes. *Transportation Research Record* 1999(1), 71-78.

Vlahogianni, E.I., 2013. Modeling Duration Of Overtaking In Two Lane Highways. *Transportation Research Part F: Traffic Psychology And Behaviour* 20, 135-146.

Wang, Q., Li, Z., Li, L., 2014. Investigation of Discretionary Lane-Change Characteristics Using Next-Generation Simulation Data Sets. *J. Intell. Transport. Syst.* 18(3), 246-253.

Washington, S., Karlaftis, M.G., Mannering, F., Anastasopoulos, P., 2020. *Statistical And Econometric Methods For Transportation Data Analysis*. CRC press.

Wu, J., Zhang, S., Singh, A.K., Qin, S., 2008. Hazard-Based Model Of Mandatory Lane

Change Duration, *17th COTA International Conference of Transportation*, China.
Yang, M., Wang, X., Quddus, M., 2019. Examining lane change gap acceptance, duration and impact using naturalistic driving data. *Transportation Research Part C: Emerging Technologies* 104, 317-331.