TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 03/10 pp471–478
DOI: 10.26599/TST.2019.9010015
Volume 25, Number 4, August 2020

Will Crash Experience Affect Driver's Behavior? An Observation and Analysis on Time Headway Variation Before and After a Traffic Crash

Yun Yue, Zi Yang, Xin Pei*, Hongxin Chen, Chao Song, and Danya Yao

Abstract: Research into the impact of road accidents on drivers is essential to effective post-crash interventions. However, due to limited data and resources, the current research focus is mainly on those who have suffered severe injuries. In this paper, we propose a novel approach to examining the impact that being involved in a crash has on drivers by using traffic surveillance data. In traffic video surveillance systems, the locations of vehicles at different moments in time are captured and their headway, which is an important indicator of driving behavior, can be calculated from this information. It was found that there was a sudden increase in headway when drivers return to the road after being involved in a crash, but that the headway returned to its pre-crash level over time. We further analyzed the duration of the decay using a Cox proportional hazards regression model, which revealed many significant factors (related to the driver, vehicle, and nature of the accident) behind the survival time of the increased headway. Our approach is able to reveal the crash impact on drivers in a convenient and economical way. It can enhance the understanding of the impact of a crash on drivers, and help to devise more effective re-education programs and other interventions to encourage drivers who are involved in crashes to drive more safely in the future.

Key words: crash impact; driving behavior; time headway; survival time; Cox proportional hazards model

1 Introduction

Globally, road traffic accidents cause 1.25 million deaths and over 50 million injuries every year[1]. Crashes have an economic cost of over US $500 billion per annum, not to mention the physical and psychological effects that continue to haunt the individuals involved after the event. At present, practitioners and researchers in the field pay most attention to the rehabilitation of severely injured victims, especially those suffering from post-traumatic stress disorder[2, 3]. Meanwhile, drivers involved in small accidents, which are much more frequent than severe ones, may also be affected in terms of their driving behavior when they return to the road. Perhaps due to limited data, this impact is rarely mentioned in the literature. In this study, we draw on a new source of data collected in a city in northern China, trying to answer the following questions about the impact on drivers of being involved in a road accident:

(1) Does the experience of a crash influence a driver's subsequent driving behavior?

(2) Do drivers exercise more care after a crash?

(3) What factors lie behind any changes in driving behavior after a crash?

Answers to these research questions may help us reveal the impact of crashes from a novel perspective, and also help with programs for crash prevention and strategies to encourage safe driving.

To measure post-crash driving behavioral change we use time headway, which is widely used to describe driving behavior in safety performance evaluations

• Yun Yue, Zi Yang, Xin Pei, Hongxin Chen, and Danya Yao are with Department of Automation, Tsinghua University, Beijing 100084, China. E-mail: ziyueyunxiao@163.com; yangzibruce@foxmail.com; peixin@tsinghua.edu.cn; chenhongxin@tsinghua.edu.cn; yaody@tsinghua.edu.cn.
• Chao Song is with National Engineering Laboratory for Public Safety Risk Perception and Control by Big Data (PSRPC), China Academy of Electronics and Information Technology, Beijing 100041, China. E-mail: songch35@mail.ustc.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2019-03-17; revised: 2019-04-04; accepted: 2019-04-10

in both research[4] and law enforcement in many countries, including the US, Germany, and Sweden. Previous studies have suggested that a shorter time headway is always present alongside other risky driving behaviors[5,6], and is significantly associated with a higher crash risk[7,8]. Headway data has generally been collected through experiments on driving simulators or with the use of on-board units. Studies based on these data have helped to analyze the role of headway to some extent, but the reliability of the results may be undermined by the limited number of drivers and short experiment durations. In this study, we extract time headway data using the output of a traffic video surveillance system, in which cameras are deployed at intersections and on road segments to capture the plate information of passing vehicles. The time headway information is calculated from the continuous recordings of vehicles in procession. When this information is put together with crash data containing the crash time and vehicle plate information, the corresponding driver's time headway before and after a crash can be plotted and analyzed.

After observing any change in time headway before and after a crash, we analyze the factors behind changes in driving behavior. Reviewing the time headway variations of most drivers reveals a typical pattern: the time headway is around the baseline of each driver before a crash; on returning to the road after a crash it initially increases to a higher level, then returns back to the pre-crash baseline after several days of driving. We therefore adopt the survival time of the increased headway as a measure of post-crash driving behavior change. The survival time can be neatly fitted into a Cox-regression survival model. This survival analysis model was first proposed in the fields of biology and medicine to model the survival time of viruses, organisms, or patients; it has now become a widely used method in the field of road safety to model crash occurrence times[9], response time[10], clearance time[11], and levels of driving risk[12–15]. We adopt this model to conduct the survival analysis for changes in driving behavior after a crash. It can not only predict the duration of the survival time, but also analyze the relationship between the survival time and other factors, including the characteristics of the driver, the vehicles involved, and the nature of the crash.

The remainder of this paper is organized as follows: Section 2 describes the collection and processing of the data; Section 3 presents the specification of the Cox-regression model for survival analysis; Section 4 discusses the results and findings; and Section 5 concludes the paper with a summary of findings, implications, and study limitations.

2 Data Collection and Processing

The city selected for the study is located in northern China and has a population of approximately 4.5 million people living in an area of about 50 km^2. The central city area contains about 60 major roads and 76 intersections and, according to the traffic surveillance system data, there are over 100 000 vehicles on these roads at a given time. We extract the time headway as a proxy measure for driving behavior, and the time headway survival time for drivers involved in an accident is calculated. A comprehensive database made up of driving behavior, survival time, and related factors is established from the data collection.

2.1 Time headway

Time headway data is obtained from a traffic surveillance system with high-definition cameras deployed at important intersections and road segments to capture detailed information on each passing vehicle in a typical city in northern China. Data are collected for the 2014 calendar year, from 1 January to 31 December. A total number of 84 cameras have been installed at major intersections throughout the city. Figure 1 shows the cameras in the central city area. From these cameras, the traffic surveillance system can extract detailed information on each passing vehicle, including the vehicle type, license plate, speed, traveling lane, and

Fig. 1 Arterial road network with monitoring locations at major intersections.

passing time. A total of 606 million records were taken during 2014. Based on these continuous records of the procession of vehicles, the time headway for each vehicle passing by the monitoring points could be calculated.

Crash records during the study period are also obtained from the police, including the date and time of each accident, the crash location, the license plates of the vehicles involved, and personal information on the drivers, such as gender, age, and years of driving experience. The traffic surveillance data and crash records are then matched to filter for the vehicles and drivers involved in accidents. For each vehicle involved in one or more crashes, the average daily time headway for each date the driver travelled is ordered chronologically to form a driving behavior dataset. Records with a time headway of no more than five seconds are further extracted, according to the Highway Capacity Manual definition of the time headway for a following car[16]. The dataset is further smoothed by applying a window time of seven days to minimize noise. Due to the truncation of data in January and December, the observation period is set as February 2014 to November 2014 instead of the whole calendar year. As some vehicles were written off in accidents and thus no longer exposed on the road, we only include vehicles that had traveling data within 30 days of the crash occurrence. The final dataset contains 234 drivers, which is regarded as sufficient to build a Cox survival model.

2.2 Survival time

In Fig. 2 we give an example to show the definition of survival time. Figure 2 presents a typical pattern of driving behavior variation after a crash. Day 0 is the day when the crash occurred, before which the driving behavior in terms of time headway is measured at 2.1 seconds. After the crash, this time headway suddenly increased, but it returned to 2.1 seconds after 48 days. The duration of the increase (48 days in this case) is defined as the survival time.

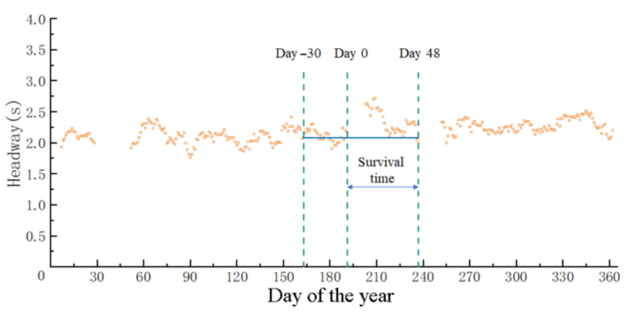


Fig. 2 A typical process of driving behavior variation with a crash happened on Day 0.

Due to the limited observation period, there are two types of survival data available to the survival analysis. Type 1 data items are complete, meaning that they encapsulate a complete transition with observations of both a triggering event and an ending event. In this study, the triggering event is the crash and the ending event is the post-crash average daily time headway of the vehicle returning to its pre-crash level. Type 2 data items are censored or, more specifically, right-censored data in our study. Censoring refers to the condition in which some part of the information of a complete transition has not been observed. Left-censored data arises from a cessation in data collection when the triggering event (i.e. a road crash) has not occurred yet; these occurrences had already been cleaned from our data. Right-censored data does show the occurrence of the triggering event, but no ending event is observed within the observation window. In this study, that means that we do not have access to the data for when the time headway returned to its pre-crash level. In our database, among the 234 qualified profiles of crash-related drivers, we have 222 Type 1 drivers and 12 Type 2 drivers.

2.3 Related factors

In order to reveal the factors behind the variations in survival time for different drivers, we consider three major categories of potential factors that may be related to the impact that a crash has on driving behavior, as follows:

(1) Driver-related factors, made up of sex, age, driving experience, occupation type, average daily time headway before the crash, and driving frequency before the crash (number of days travelled as a percentage of the total number of natural days);

(2) Crash-related factors, made up of whether the driver was injured, the number of other people injured, crash liability, any direct property loss caused by crash, and the crash type;

(3) Vehicle-related factors, made up of vehicle ownership (private or commercial).

These variables can be further classified into (1) numerical variables (age, driving experience, direct property loss caused by crash, average daily time headway before the crash, driving frequency before the crash, and number of other people injured) and (2) categorical variables (sex, occupation type, whether the driver was injured, crash type, and vehicle ownership).

A survival analysis is adopted to establish a risk model and analyze the relationships between changes in driver behavior and potential factors. Before placing the related factors into the survival model as independent variables, a correlation test is conducted for each pair of possible factors. As there is a significant correlation between age and driving experience, only the latter is considered in the model. Table 1 presents the statistical characteristics of the independent variables.

3 Method

3.1 Survival analysis

A survival analysis considers both the event outcome (either survival or death) of the observed object and the length of its survival time. This approach can take advantage of the temporal variations in probability and can offer point-in-time predictions. In our study, the event is defined as the occurrence of a crash, and the survival time is defined as the number of natural days after a crash until the driver's average daily time headway returns to its pre-crash levels.

Three fundamental functions in survival analysis are the probability density function, survival function, and hazard function. The probability density function $f(t)$ represents the probability that the impact of a traffic crash on the driver's driving behavior ends in a $[t, t+\Delta t]$ period, and is defined as

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T < t + \Delta t)}{\Delta t} \quad (1)$$

The survival function $S(t)$ describes the progression of the survival rate over time. It represents the probability that a traffic crash affects the driver's driving behavior for longer than a given time t, and is defined as

$$S(t) = P(T \geqslant t) \quad (2)$$

The hazard function $\lambda(t)$ refers to the probability that a crash stops influencing the driver in a $[t, t+\Delta t]$ period, given that the crash still influences the driver at time t. The hazard function is therefore a conditional probability

Table 1 Statistical summary of the observations incorporated into the proposed model, where the number of observations is 234.

Variable	Min	Max	Mean	Standard deviation	Count	Proportion (%)
Survival time (day)	9	265	44.94	40.38		
Driving age (year)	1	32	8.35	6.29		
Pre-crash exposure frequency (%)	3.39	99.44	69.02	26.61		
Pre-crash headway (s)	1	4	2.56	0.30		
Direct property loss (RMB)	0	50 000	2838.89	5895.00		
Number of people injured	0	6	1.10	0.68		
Gender						
Male					193	82.48
Female					41	17.52
Whether the driver injured						
Yes					6	2.56
No					228	97.44
Crash liability						
Full & primary responsibility					171	73.08
Secondary & no responsibility					33	14.10
Equal responsibility					30	12.82
Personnel occupation type						
Unemployed					116	49.57
Stable job employed					64	27.35
Self-employed					54	23.08
Crash type						
Multi-vehicle crash					196	83.76
Vehicle-pedestrian crash					33	14.10
Single-vehicle crash					5	2.14
Vehicle ownership						
Private car					182	77.78
Commercial vehicle					49	20.94
Others					3	1.28

function, and is defined as

$$\lambda(t)=\lim_{\Delta t\to 0}\frac{P(t\leqslant T<t+\Delta t|T\geqslant t)}{\Delta t}=\lambda_0(t)g(X) \tag{3}$$

where λ_0 denotes the baseline of the hazard rate (a function of time t and independent of the covariate X), X is the vector of influencing factors, and $g(X)$ is the log-linear form of X. Each of these functions can be derived from the other two with a relationship of Eq. (4):

$$\lambda(t)=\frac{f(t)}{S(t)} \tag{4}$$

The statistical methods for survival analysis are divided into non-parametric models, parametric models, and semi-parametric models. In the original analysis, the non-parametric model estimates the survival rate from simple statistics of the samples in the dataset, regardless of their distribution. This approach is easy to implement but the results may not provide enough insights as the data are not fully analyzed. One commonly used model of this type is the Kaplan-Meier estimator model[17], defined as

$$S(t)=\prod_{t_i<t}\frac{n_i-d_i}{n_i} \tag{5}$$

where d_i is the number of samples with events ending by time t and n_i is the number of samples with events ending by t_i. The parametric method assumes that the survival time is influenced by certain parameters that follow specific distributions, such as exponential distribution or Weibull distribution. The survival time can then be further analyzed according to the characteristics of the known distribution. The semi-parametric regression model combines the advantages of both the parametric and non-parametric method, and is usually adopted when the distribution type of the survival time is unknown. This approach is also very helpful to analyze the probable influences of multiple factors on survival time and survival rate. One typical semi-parametric method is the Cox proportional hazards regression model[18] .

3.2 Cox proportional hazards regression model

By introducing time into the model, the Cox regression model is more flexible and able to explore more information from multiple factors and more effectively conduct multivariate prognostics. Cox regression assumes that the factors multiplicatively affect the hazard function regardless of the form of the hazard function. This gives it a great advantage when solving problems with unknown hazard functions, which is the case in this study. In addition, Cox regression utilizes the rank ordering of survival time to analyze the effects of different factors, and thus the results are less affected by abnormal samples[19].

Typically, in a Cox regression model, the hazard function of the survival time is expressed as

$$\lambda(t|X)=\frac{f(t|X)}{S(t|X)} \tag{6}$$

where X represents the covariates (factors or predictors), $f(t|X)$ is the density function of the survival time T, and $S(t|X)$ is the survival function. In this study, the covariate X is a constant vector composed of several time-invariant factors; thus, the ratio $\lambda(t|X_1)/\lambda(t|X_2)$ is constant over t if $X_1\neq X_2$. According to Eq. (3), the Cox regression model can be further expressed as

$$\lambda(t|X)=\lambda_0(t)\exp(\beta_1x_1+\cdots+\beta_nx_n) \tag{7}$$

where β is a coefficient vector representing the effects of factors on the hazard, and $\exp(\beta_i)$ is defined as the hazard ratio, which indicates that when the factor x_i increases by one unit, the hazard probability increases by $\exp(\beta_i)$ and the survival time becomes shorter accordingly. The exponential expression is a function of covariance X and is independent of time.

4 Results and Discussion

A Cox proportional hazards regression model is established to model the effect of potential factors on survival time. Maximum-likelihood-based probabilistic statistical methods are used to estimate the model parameter of β, which indicates the positive or negative influence of the corresponding factor on the hazard and the survival time. If $\beta_i>0$ at the 5% level of significance then the corresponding factor x_i affects the hazard positively (thus reduces the survival time), whereas if $\beta_i<0$ at a 5% significant level then the corresponding factor x_i affects the hazard negatively (thus extends the survival time). We also calculate the hazard ratio of factor x_i as $\exp(\beta_i)$.

Table 2 shows the results of the Cox regression model applied to our dataset with point estimation of model parameters, hazard ratio, and 95% Confidence Interval (CI) scores. In order to assess the goodness of fit of the proposed model, a traditional Likelihood Radio (LR) test is conducted. The LR-statistic is a Chi-square distributed with a degree of freedom equal to the number of independent variables. According to the results shown in Table 2, the LR-statistic is significant at 1% level and therefore the overall goodness-of-fit of the proposed

Table 2 Survival analysis results of Cox proportional hazards model, where * represents statistically significant at the 5% level, **LR=1913 (Chi=49, *p*-value<0.01)**.

Variable	β_i	$\exp(\beta_i)$	95% CI score
Driving age (year)	0.01	1.01	$(-0.14, 0.16)$
Pre-crash headway (s)	0.82*	2.27	$(0.26, 1.38)$
Pre-crash exposure frequency	1.23*	3.44	$(0.67, 1.796)$
Direct property loss (RMB)	−0.46*	0.63	$(-0.80, -0.12)$
Number of people injured	−0.07	0.93	$(-0.28, 0.13)$
Gender			
Male	(Control)		
Female	0.28	1.33	$(-0.10, 0.67)$
Whether the driver injured			
Yes	(Control)		
No	−0.27	0.77	$(-1.34, 0.80)$
Crash liability			
Full & primary responsibility	0.13	1.13	$(-0.31, 0.56)$
Secondary & no responsibility	0.09	1.09	$(-0.53, 0.70)$
Equal responsibility	(Control)		
Personnel occupation type			
Unemployed	0.44*	1.55	$(0.06, 0.82)$
Stable job employed	0.24	1.27	$(-0.17, 0.65)$
Self-employed	(Control)		
Crash type			
Multi-vehicle crash	−0.79	0.45	$(-1.93, 0.35)$
Vehicle-pedestrian crash	−1.25*	0.29	$(-2.44, -0.06)$
Single-vehicle crash	(Control)		
Vehicle ownership			
Private car	−0.42*	0.66	$(-0.78, -0.05)$
Commercial vehicle	(Control)		
Others	0.09	1.10	$(-1.12, 1.31)$

model is significant and acceptable.

Pre-crash driving exposure frequency, pre-crash headway, and an occupation of unemployed all have $\beta_i > 0$ at a 5% significance level and are therefore positively associated with the hazard, which represents the probability of crash impacts disappearing. Therefore, these factors reduce the survival time significantly. With respect to pre-crash driving frequency ($\beta_i = 1.23$), it seems that the more frequently the drivers travel, the less a crash influences their subsequent driving behavior. This result shows that the driving habits of regular drivers may be more persistent than those of irregular drivers. Thus, their time headway returns to its pre-crash level more quickly, i.e., the impact of a crash has a shorter duration. With respect to pre-crash headway within one month before the crash ($\beta_i = 0.82$), the longer the pre-crash headway, the more quickly the post-crash headway returns to its pre-crash level. It is obvious that with a higher baseline, although the headway after a crash is increased, it will return to its previous level more quickly. In other words, the behavior of drivers with a longer pre-crash time headway is less affected by a crash as they are already cautious drivers. With respect to occupation type, in contrast to self-employed drivers engaging in some form of private business, drivers without stable employment experience a shorter duration of crash impact ($\beta_i = 0.44$). This may reflect the different educational backgrounds of the drivers, and probably different levels of economic distress.

Direct property loss, crash type, and vehicle ownership are negatively related to the hazard, with ($\beta_i < 0$) at a 5% significance level, therefore increasing the survival time. With respect to the direct property loss factor ($\beta_i = -0.46$), the greater the property loss suffered in the crash, the longer the survival time. This is in line with our intuition that a greater loss of property will have more impact on the driver, and thus it will take longer for the driver to return to his or her pre-crash driving behavior. With respect to crash type, when compared to single-vehicle crashes, a vehicle-pedestrian crash ($\beta_i = -1.25$) will have more impact on the driver in terms of longer survival time with a longer time headway. This indicates that when a pedestrian is injured in a traffic accident, the driver's driving behavior will take a long time to return to its pre-crash level, and may even never do so. Again, this is an intuitive result, as we would expect that hitting a pedestrian would have an major impact on the mentality of a driver. With respect to vehicle ownership, when compared to drivers of commercial vehicles, the change in driving behavior of private car drivers ($\beta_i = -0.42$) has a longer survival time with an increased time headway. This can be interpreted in line with the different patterns of driving behavior and vehicle use arising from the different ownership types.

5 Conclusion

In this paper, we proposed a novel approach to examine the impact of a road accidents on drivers with the use of traffic surveillance data. Time headway is used as the indicative measure of driving behavior. Most drivers respond to crashes by driving more carefully (with a longer time headway) for a certain amount of time before eventually returning to their pre-crash driving behaviors. The duration until the headway returns to its pre-crash level is modeled using a

Cox regression survival model. The results show that there are many factors influencing this duration: pre-crash driving frequency, pre-crash driving behavior, unemployment, direct property loss suffered in the crash, crash type (especially the involvement of a pedestrian), and vehicle ownership type (commercial or private). This information will help us better understand the impacts of road accidents on drivers, and to devise more effective re-education programs and other interventions to encourage drivers who have been involved in crashes to drive more safely in the future.

The method we proposed in this paper is transferable to other cities where the traffic surveillance data is available and validations of the study conducted in other cities would be welcome. Due to limitations in the data and approach, there is much follow-up research to be done before we can make optimal use of the results of this study. First, we only use the time headway to represent driving characteristics, but there are many other aspects of driving behavior that are important to safety. Results taking these other aspects into consideration would be informative. Second, the duration of change in headway shows the existence of a crash impact decay, but this study did not consider the magnitude of headways due to sample size and the over-diversified background of the drivers in the sample. This information could also be useful. Finally, the results in this study show the importance of different factors in crash impact, but there is still a large step to take from these results to the design and implementation of effective interventions. Nonetheless, we hope that the results of the study are helpful to professionals in traffic management, psychology, and policy making who are dedicated to encouraging safe driving.

Acknowledgment

The work was supported by the National Natural Science Foundation of China (No. 71671100), the Joint Research Scheme of the National Natural Science Foundation of China/Research Grants Council of Hong Kong (Nos. 71561167001 and N_HKU707), the Director Foundation Project of National Engineering Laboratory for Public Safety Risk Perception and Control by Big Data (PSRPC), and the Research Funds of Tsinghua University (No. 20151080412).

References

[1] World Health Organization, Global status report on road safety, https://www.who.int/violence_injury_prevention/road_safety_status/2015/en/, 2015.

[2] S. J. Fredman, J. G. Beck, P. Shnaider, Y. Y. Le, N. D. Pukay-Martin, K. Z. Pentel, C. M. Monson, N. M. Simon, and L. Marques, Longitudinal associations between PTSD symptoms and dyadic conflict communication following a severe motor vehicle accident, *Behav. Ther.*, vol. 48, no. 2, pp. 235–246, 2017.

[3] M. Heron-Delaney, J. Kenardy, E. Charlton, and Y. Matsuoka, A systematic review of predictors of posttraumatic stress disorder (PTSD) for adult road traffic crash survivors, *Injury*, vol. 44, no. 11, pp. 1413–1422, 2013.

[4] K. Vogel, A comparison of headway and time to collision as safety indicators, *Accid. Anal. Prev.*, vol. 35, no. 3, pp. 427–433, 2003.

[5] J. Y. Duan, Z. Z. Li, and G. Salvendy, Automatic imitation of risky behavior: A study of simulated driving in China, *Traffic Inj. Prev.*, vol. 13, no. 5, pp. 442–449, 2012.

[6] X. B. Qu, Y. Kuang, E. Oh, and S. Jin, Safety evaluation for expressways: A comparative study for macroscopic and microscopic indicators, *Traffic Inj. Prev.*, vol. 15, no. 1, pp. 89–93, 2014.

[7] H. X. Chen, X. Pei, Z. Zhang, D. Y. Yao, Q. J. Feng, and Z. Z. Wang, Driving behavior differences between crash-involved and crash-not-involved drivers using urban traffic surveillance data, in *2016 IEEE Int. Conf. on Service Operations and Logistics, and Informatics (SOLI)*, Beijing, China, 2016, pp. 249–254.

[8] L. Evans and P. Wasielewski, Do accident-involved drivers exhibit riskier everyday driving behavior? *Accid. Anal. Prev.*, vol. 14, no. 1, pp. 57–64, 1982.

[9] A. Vadeby, Å. Forsman, G. Kecklund, T. Åkerstedt, D. Sandberg, and A. Anund, Sleepiness and prediction of driver impairment in simulator studies using a Cox proportional hazard approach, *Accid. Anal. Prev.*, vol. 42, no. 3, pp. 835–841, 2010.

[10] E. Lee, J. Wu, T. Kang, and M. Craig, Estimate of mortality reduction with implementation of advanced automatic collision notification, *Traffic Inj. Prev.*, vol. 18, no. sup1, pp. S24–S30, 2017.

[11] J. T. Lee and J. Fazio, Influential factors in freeway crash response and clearance times by emergency management services in peak periods, *Traffic Inj. Prev.*, vol. 6, no. 4, pp. 331–339, 2005.

[12] S. Lawpoolsri, J. Y. Li, and E. R. Braver, Do speeding tickets reduce the likelihood of receiving subsequent speeding tickets? A longitudinal study of speeding violators in Maryland, *Traffic Inj. Prev.*, vol. 8, no. 1, pp. 26–34, 2007.

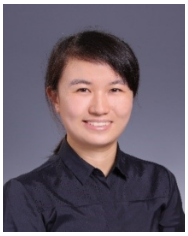
[13] R. Purssell, D. Brown, J. R. Brubacher, J. Wilson, M. Fang, M. Schulzer, E. Mak, R. B. Abu-Laban, R. Simons, and T. Walker, Proportion of injured drivers presenting to a tertiary care emergency department who engage in future impaired driving activities, *Traffic Inj. Prev.*, vol. 11, no. 1, pp. 35–42, 2010.

[14] R. A. Purssell, H. Chan, D. Brown, M. Fang, and J. R. Brubacher, Risk factors for subsequent impaired driving by injured passengers, *Traffic Inj. Prev.*, vol. 15, no. 4, pp. 355–360, 2014.

[15] C. Roque and M. Jalayer, Improving roadside design policies for safety enhancement using hazard-based duration modeling, *Accid. Anal. Prev.*, vol. 120, pp. 165–173, 2018.
[16] Transportation Research Board (TRB), *Highway Capacity Manual.* 5th ed. Washington, DC, USA: National Research Council, 2010.
[17] E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.
[18] D. R. Cox, Regression models and life-tables, *J. Royal Statist. Soc. B*, vol. 34, no. 2, pp. 187–220, 1972.
[19] F. E. Jr Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. Springer, 2015.

Yun Yue received the BS degree from Chang'an University, China, in 2004, and the MS degree from Tsinghua University, China, in 2011. She is a PhD student at Tsinghua University, studying data-driven traffic safety analysis.

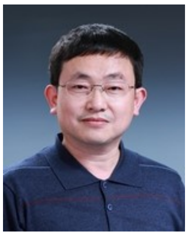
Zi Yang received the BS degree from Sun Yatsen University in 2007, the MS degree from the Chinese University of Hong Kong in 2011, and the PhD degree from the University of Hong Kong in 2017. He is currently a Post-doc fellow in the Department of Automation, Tsinghua University. His research interests include road safety evaluation, traffic safety simulation, and travel behavior analysis.



Xin Pei received the BS and MS degrees from Tsinghua University, China, in 2005 and 2007, respectively, and the PhD degree from the University of Hong Kong in 2011. She is currently a research associate professor with Department of Automation, Tsinghua University. Her current research interests include road safety evaluation and driving behavior analysis. She is Principal Investigator (PI)/co-PI of 3 road safety related NSFC projects. She has published more than 50 SCI/SSCI/EI indexed papers, especially, there are 7 papers published on *Accident Analysis and Prevention*, which is a top journal for road safety analysis.

Hongxin Chen received the BS and PhD degrees from Tsinghua University, China in 2010 and 2016, respectively. He is currently a lecturer at Tsinghua University. His current research interests include road traffic safety, driving behavior analysis, and ideological and political education.

Chao Song received the bachelor degree from Nanjing Tech University in 2012, and the PhD degree from University of Science and Technology of China in 2017. He is currently working at China Academy of Electronics and Information Technology as a system engineer. His research interests include emergency decision-making, geography information system, and public safety.



Danya Yao is currently a professor at Tsinghua University. He received the BEng, MEng, and PhD degrees from Tsinghua Unversity in 1988, 1990, and 1994, respectively. His active research areas include vehicle-infrastructure cooperation systems, advanced detection technology, and systems engineering. He has published more than 100 papers. He is PI/co-PI of more than 5 national research programs including 863 Program, 973 Program, and NSFC Project. He was ever the chief expert of the National High-Tech Research and Development Program (863) Research of China on key technologies of intelligent vehicle-infrastructure co-operation system.