Driving behavior differences between crash-involved and crash-not-involved drivers using urban traffic surveillance data

Chen Hongxin
Department of Automation,
Tsinghua University
Beijing, China
chen-hx06@mails.tsinghua.edu.cn

Pei Xin
Department of Automation,
Tsinghua University
Beijing, China
peixin@tsinghua.edu.cn

Zhang Zuo
Department of Automation,
Tsinghua University
Beijing, China
zhangzuo@tsinghua.edu.cn

Yao Danya
Department of Automation,
Tsinghua University
Beijing, China
yaody@tsinghua.edu.cn

Feng Qiaojun
Department of Automation,
Tsinghua University
Beijing, China
fqj13@mails.tsinghua.edu.cn

Wang Zizhuo
Department of Automation,
Tsinghua University
Beijing, China
nmwangzizhuo@163.com

***Abstract*—With technology such as in-vehicle data collection systems, driving data including mileage, speed, acceleration can be collected and analyzed by many researchers. However, in these studies, data could be collected only from a few selected drivers. In addition, drivers knowing that they were participating experiments might drive differently from natural. Furthermore, few researches took advantage of headway, which requires data from not only objective vehicles but also vehicles nearby. Many urban traffic surveillance systems built in recent years have brought new opportunities for researches. In this paper, urban traffic surveillance data at both intersections and road segments were used, so that data of numerous vehicles including objective vehicles and vehicles nearby could be collected, and indicators such as headway of vehicles could be calculated. The differences of driving behavior between crash-involved and crash-not-involved drivers were then analyzed. It was found that crash-involved drivers tended to keep less headways than crash-not-involved drivers when driving through intersections in everyday driving behavior. In the days before the crashes, this tendency of male drivers was stronger than female drivers. For road segments, compared with crash-not-involved drivers, crash-involved drivers' headways were seen less, and crash-involved drivers' speeds under free flow condition were seen larger at certain time frames. The result suggests that there is a great potential to taking advantage of urban traffic surveillance data to identify at-risk drivers.**

Keywords—driving behavior; time headway; speed; crash

I. INTRODUCTION

Human failure is the most important factor that may lead to an accident [1]. Studies on driving behavior help to improve traffic safety level, and have been one of the hot topics in traffic safety.

Many previous studies focused on the relation between driving behavior parameters and the number/severity of crashes [2,3,4,5]. The driving behavior parameters include acceleration, speed, etc. Researchers have used average speed, the standard deviation of speed, the percentage of vehicles exceeding the speed limit, or the ratio of standard deviation to mean speed, to analyze driving behavior [6].

The limitation of these studies were they concentrated on certain locations or at a specified short distance of roadway (location-based studies), which couldn't represent the behavior of everyday driving. In addition, they only studied the relation between parameters and crashes at macroscopic level. The individual driving behaviors were ignored. Moreover, previous studies didn't statistically evaluate the differences of driving behavior between crash-involved and crash-not-involved drivers [7].

In recent years, with the increasing adoption of in-vehicle data recorders, there were more and more studies (vehicle-based studies) taking advantage of everyday driving data to analyze driving behavior [8,9,10,11]. Several studies had been

978-1-5090-2927-3/16/$31.00 ©2016 IEEE

done to reveal the relationship between mileage, velocity, acceleration, critical driving events and accident involvement. However, due to the limitation of in-vehicle data recorders, records only contained information of objective vehicles. It is difficult to obtain parameters such as headway which requires data from not only objective vehicles but also vehicles nearby, with many current in-vehicle equipment.

Many urban traffic surveillance systems have been built in these years. With these systems, data of individual drivers with and without crash involvements could be obtained.

Compared with previous location-based studies, in this study, data could be obtained from most intersections and road segments, drivers' everyday driving behavior could be revealed, and the differences of driving behavior could be analyzed. Compared with vehicle-based studies, with urban traffic surveillance systems, data from both objective vehicles and vehicles nearby could be obtained, so that indicators such as headways could be analyzed.

As one of the indicators, headway could be used to estimate the criticality of a certain traffic situation [12]. In some countries, this indicator is also used to impose fines for close following. However, few researchers analyzed the differences of headways between crash-involved drivers and crash-not-involved drivers.

Therefore, this paper collected driving data from urban traffic surveillance systems and analyzed the differences of driving behavior parameters including headway and speed, between crash-involved drivers and crash-not-involved drivers. The result suggests that there is a great potential to taking advantage of urban traffic surveillance data to identify at-risk drivers.

The paper is organized as follows. In Section 2, we introduce the urban traffic surveillance data. In Section 3 the statistical methodologies used in this paper is presented. In Section 4, we present the results of differences of driving behavior between crash-involved and crash-not-involved drivers and, finally, in Section 5 we conclude.

II. DATA COLLECTION

This section introduces the urban traffic surveillance data analyzed in this paper.

This study collected 2 cities' traffic surveillance data. City A located in north China, with a population of about 4.5 million. The other city B located in east China, with a population of about 3.6 million. There were 79 monitored intersections in city A, and 84 monitored intersections, 52 road segments in city B. The urban traffic surveillance systems equipped in two cities generate records when drivers driving through the intersections or road segments. Take the records of city A as an example, there were about 1.5 million driving data recorded every day.

The records contained time, location, license plate, vehicle type, speed and so on. Data of city A in the year 2014 and of city B from October 2015 to April 2016 was used in this study. At the same time, this study also obtained crash information of these two cities during the same period that driving data recorded. Crash information contained crash time, crash location, crash severity, crash-involved license plate, drivers' gender, drivers' age, drivers' driving experience and so on.

III. DATA PROCESSING AND STATISTICAL METHODOLOGIES

This section introduces process of data selection. Meanwhile analysis of variance (ANOVA) and paired-sample t-test were introduced in this section.

There were three steps of crash-involved drivers' selection. First of all, because of the differences of length of vehicles between large vehicles and light-duty vehicles, the headway contributions vary a lot between two groups of vehicles. Thus in this study, light-duty vehicles were selected as research objects. Secondly, due to the unfamiliar with the road condition, drivers might perform different driving habits. Thus, vehicles from other cities were not taken into account in this study. Lastly, vehicles for special applications such as police vehicles, learner-driven vehicles, also shouldn't be taken into account.

In city A, there were 1015 crashes happened in the year 2014, and there were 784 drivers selected as the research objects. Among these crash-involved drivers, there were 675 male drivers and 109 female drivers. In city B, there were 341 drivers selected as the research objects and there were 301 male drivers and 40 female drivers.

Since drivers might adjust their driving habits after the crash, thus this study selected data from the day when crashes happened and several days before as the crash-involved drivers' data.

Comparisons of headways and speeds under different situations might lead to incorrect conclusion. In order to minimize the impact of external environment, this study selected data following the crash-involved drivers' data as the reference data. The external environment of the reference data were similar with that of the crash-involved drivers' data, as they were close in time and space. During the observation period, less than 0.06% of the light-duty drivers had been involved in crashes. Due to the scarcity of the traffic crashes, reference drivers were unlikely to get involved in crashes. This study examined the reference drivers, and none of the reference drivers had been involved in crash during the observation period. Therefore, data following the crash-involved drivers' date were selected as the crash-not-involved data.

In this study, a day was divided into six different time frames: early morning time (12a.m.–6 a.m.), a.m. peak (6 a.m.–9 a.m.), morning time (9 a.m.–12 p.m.), afternoon time (12 p.m.–5 p.m.), p.m. peak (5 p.m.–8 p.m.), and nighttime (8 p.m.–12 a.m.). Since there were not enough travel data during early morning time, this research mainly focus on other five time frames.

When analyzing headways at intersections, it was important to consider the effects of traffic signal and stop line. According to the analysis in Highway Capacity Manual [13], driver of the first vehicle in the queue should observe the signal change to green and react to the change. The second vehicle follows a similar process, except that the reaction and acceleration period can occur while the first vehicle is beginning to move. Thus, the headway of one vehicle is generally less than that of the vehicle ahead at a signalized intersection.

ANOVA and paired-sample t-test were used in this study to verify differences in means of headways and speeds between two groups by using SPSS 19 for windows. Both of these two methods are heavily used in the analysis of data. ANOVA provides a statistical method of whether or not the means of groups are equal. In this study, that means whether or not the means of headways and speeds of crash-involved and crash-not-involved drivers were equal. Since this study selected the data following the crash-involved drivers' data as the crash-not-involved drivers' data, paired comparison method was more suitable for this analysis, thus paired-sample t-test was used in this study.

IV. Analytical Results

A. Differences in headways at intersections

As shown in Fig.1, the average headways of crash-involved drivers travelling through intersections were lower than drivers who were not involved in crashes. The differences of headways in all time frames significantly differed between the two driver groups at the 0.01 significance level based on both of ANOVA and paired-sample t-test. Despite headway changed throughout the day, drivers involved in crashes tended to keep a lower headway than the drivers not involved in crash. Table I shows the means and differences in headways between groups.

Although the headways of preceding vehicles are generally larger than that of the following vehicles at a signalized intersection in everyday driving, results of this paper showed that when the preceding vehicles were involved in crashes, the headways of them were less than those of the following vehicles which were crash-not-involved.

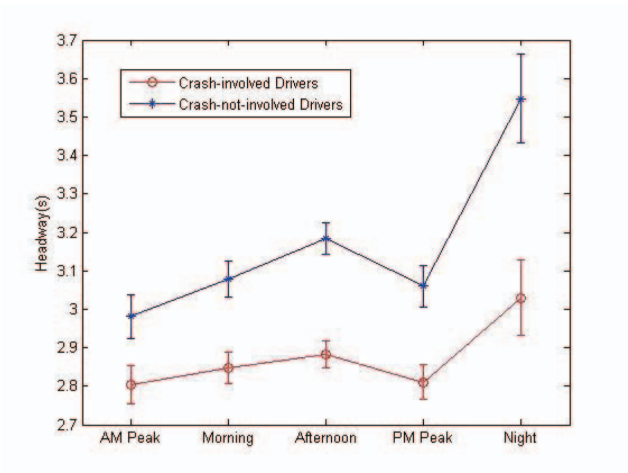

Fig. 1. Average headways at intersections by trip start times between groups.

TABLE I. Means and Differences in Headways at Intersection between Groups.

Time frames	Mean of headways (s)		Difference	Sig. (ANOVA)	Sig. (t-test)
	Drivers involved in crashes	***Drivers not involved in crashes***			
A.M. peak	2.803	2.981	-0.178	0.000	0.000
Morning*	2.847	3.079	-0.232	0.000	0.000
Afternoon	2.882	3.184	-0.302	0.000	0.000
P.M. peak	2.811	3.060	-0.249	0.000	0.000
Night*	3.030	3.547	-0.517	0.000	0.000

* Indicates a significant mean difference (α = 0.01) based on both methods.

This research took the impact of gender into consideration, and found out that both of crash-involved male drivers and female drivers performed lower headway than drivers who were not involved in crashes. Both of the differences were significant at the 0.05 significance level based on both methods (Fig.2 and Fig.3).

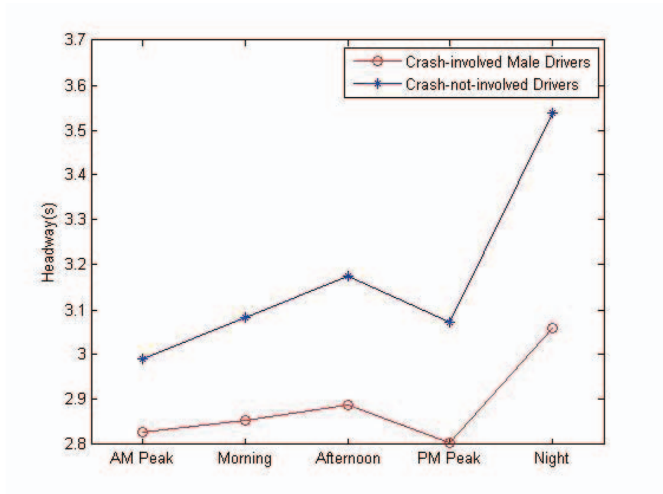

Fig. 2. Average headways at intersections by trip start times between crash-involved male drivers and crash- not-involved drivers.

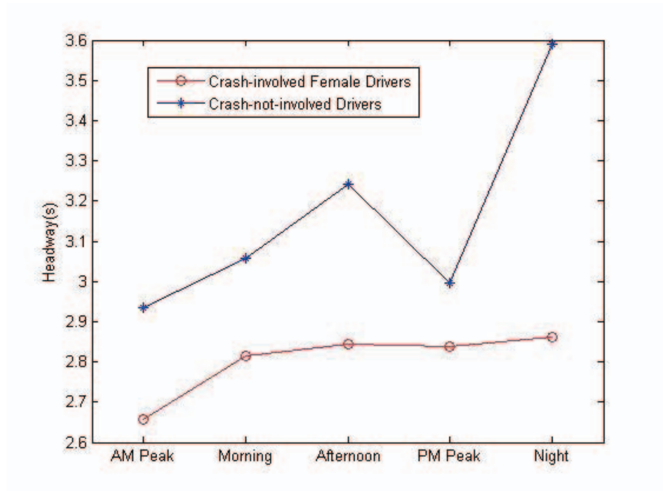

Fig. 3. Average headways at intersections by trip start times between crash-involved female drivers and crash-not-involved drivers.

In order to figure out whether the differences between crash-involved drivers and crash-not-involved drivers existed for a period of time or just in the day crash happened, this research not only analyzed the day that the crash happened, but also several days before.

As is shown in Fig.4, the average headways of crash-involved drivers travelling through intersections stayed around 2.85s, while the average headways of crash-not-involved drivers stayed around 3.13s. All the differences of 15 days significantly differed between the two driver groups at the 0.05 significance level. This result indicated that for most crash-involved drivers, keeping a low headway was a driving habit rather than a temporary change.

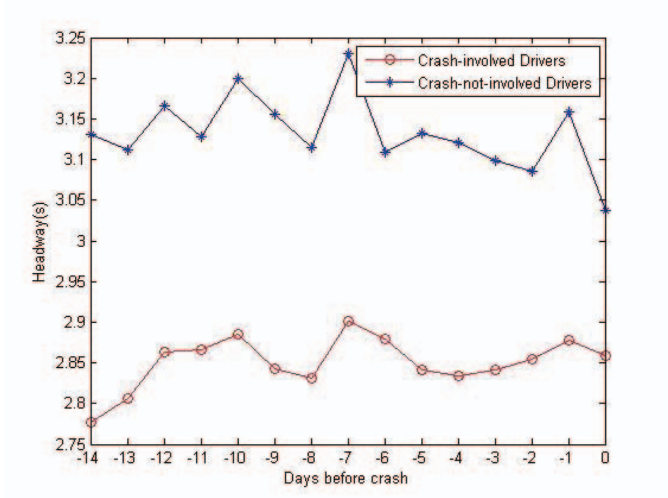

Fig. 4. Average headways at intersections by days before a crash between groups.

Similarly, the impact of gender was taken into consideration. The differences between crash-involved male drivers and drivers following them (crash-not-involved) were always exist and significant in these 15 days (Fig.5). Meanwhile, the differences between crash-involved female drivers and drivers following them (crash-not-involved) were exist and significant only in several days (Fig.6). The results suggested that compare with crash-involved male drivers, the differences between crash-involved female drivers and drivers following them (crash-not-involved) were sporadic (Table II).

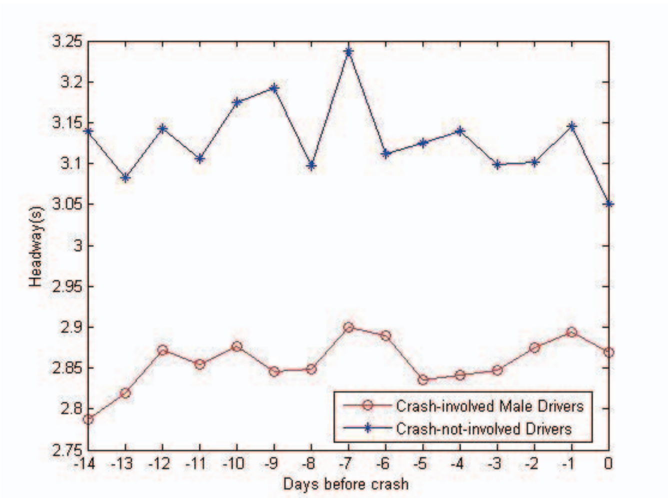

Fig. 5. Average headways at intersections by days before a crash between crash-involved male drivers and crash- not-involved drivers.

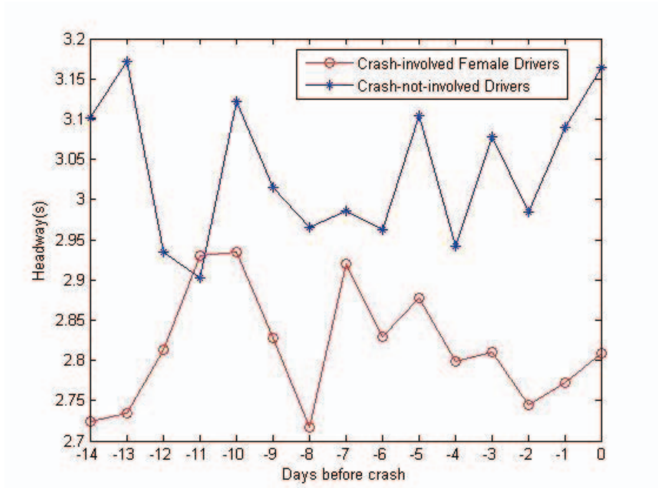

Fig. 6. Average headways at intersections by days before a crash between crash-involved female drivers and crash- not-involved drivers.

TABLE II. MEANS AND DIFFERENCES IN HEADWAYS AT INTERSECTION BASED ON GENDER AND DAYS FEFORE A CRASH BETWEEN GROUPS.

Gender	**Days**	**Mean of headways (s)**		**Difference**	**Sig. (ANOVA)**	**Sig. (t-test)**
		Drivers involved in crashes	***Drivers not involved in crashes***			
Male	-14*	2.787	3.139	-0.352	0.000	0.000
	-13*	2.820	3.083	-0.263	0.000	0.000
	-12*	2.872	3.142	-0.270	0.000	0.000
	-11*	2.855	3.106	-0.251	0.000	0.000
	-10*	2.876	3.175	-0.299	0.000	0.000
	-9*	2.845	3.192	-0.347	0.000	0.000
	-8*	2.849	3.097	-0.248	0.000	0.000
	-7*	2.899	3.237	-0.338	0.000	0.000
	-6*	2.889	3.112	-0.223	0.000	0.000
	-5*	2.835	3.126	-0.290	0.000	0.000
	-4*	2.842	3.140	-0.298	0.000	0.000
	-3*	2.847	3.098	-0.251	0.000	0.000
	-2*	2.874	3.101	-0.227	0.000	0.000
	-1*	2.894	3.145	-0.251	0.000	0.000
	0*	2.869	3.051	-0.182	0.001	0.000
Female	-14*	2.725	3.102	-0.377	0.002	0.001
	-13*	2.734	3.172	-0.438	0.001	0.000
	-12	2.813	2.935	-0.122	0.283	0.259
	-11	2.930	2.903	0.027	0.816	0.807

	-10	2.935	3.121	-0.187	0.161	0.155
	-9**	2.827	3.015	-0.187	0.102	0.070
	-8*	2.717	2.965	-0.248	0.035	0.030
	-7	2.920	2.986	-0.066	0.575	0.546
	-6	2.829	2.963	-0.134	0.249	0.198
	-5**	2.878	3.104	-0.226	0.061	0.028
	-4	2.799	2.942	-0.143	0.234	0.192
	-3*	2.811	3.078	-0.267	0.018	0.009
	-2*	2.745	2.984	-0.239	0.041	0.014
	-1*	2.772	3.089	-0.317	0.010	0.009
	0*	2.809	3.164	-0.356	0.003	0.001

* Indicates a significant mean difference (α = 0.05) based on both methods.

** Indicates a significant mean difference (α = 0.05) based on paired-sample t-test.

This research further analyzed the differences of headways at different speeds (Fig.7). At the speed under 20km/h, the differences between two groups (headway of crash-involved-drivers minus headway of crash-not-involved drivers) were positive, which indicated that at these speeds, crash-involved drivers performed higher headways than the following drivers. Whereas, at the speed above 20km/h, the differences of two groups turned out to be negative. This result might indicate that higher speeds and lower headways at intersections lead to more crashes.

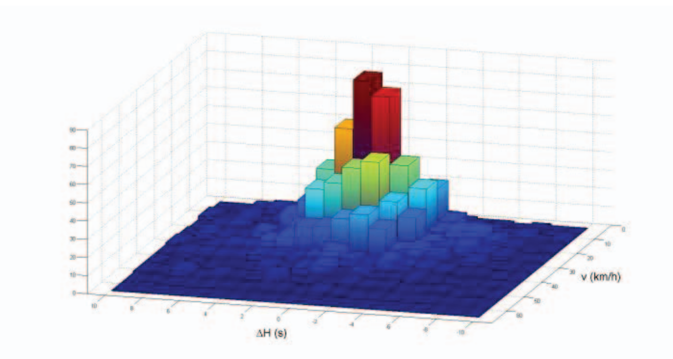

Fig. 7. Distuibution of headways differences between groups at different speeds.

B. Differences in headways at road segments

As shown in Fig.8, the average headways of crash-involved drivers traveling through road segments were generally less than crash-not-involved drivers. However, only two time frames, during morning and night, showed a significant difference based on ANOVA and paired-sample t-test (Table III).

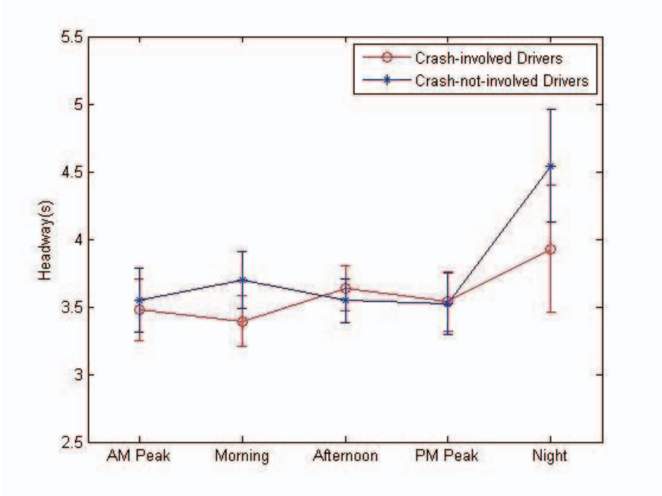

Fig. 8. Average headways at road segments by trip start times between groups.

TABLE III. MEANS AND DIFFERENCES IN HEADWAYS AT ROAD SEGMENTS BETWEEN GROUPS.

Time frames	Mean of headways (s)		Difference	Sig. (ANOVA)	Sig. (t-test)
	Drivers involved in crashes	*Drivers not involved in crashes*			
A.M. peak	3.475	3.550	-0.074	0.638	0.631
Morning*	3.394	3.696	-0.302	0.029	0.013
Afternoon	3.638	3.546	0.092	0.351	0.342
P.M. peak	3.538	3.520	0.018	0.816	0.810
Night*	3.930	4.541	-0.611	0.026	0.021

* Indicates a significant mean difference (α = 0.05) based on both methods.

C. Differences in speeds at road segments under free flow situation

It should be noted that under car-following situation, drivers tend to drive at the speed same as the premier vehicle. Thus, it is more meaningful to analyze crash-involved drivers' speed under free flow situation rather than car-following situation.

This research selected driving data which headway was above 8s and speed is above 10km/h (Table IV). Fig. 9 shows the average speeds between two groups based on trip start times. The average speed of crash-involved drivers during a.m. peak hours was 41.9km/h and that of crash-not-involved drivers was 38.9km/h, showing a difference of 3km/h. The average speed of crash-involved drivers during night hours was 42.8km/h and that of crash-not-involved drivers was 40.6km/h, showing a difference of 2.2km/h.

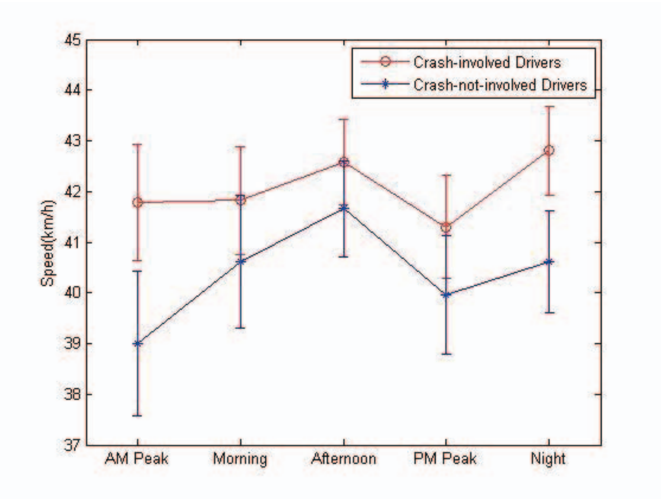

Fig. 9. Average speeds at road segments under free flow situaiton by trip start times between groups.

TABLE IV. MEANS AND DIFFERENCES IN SPEEDS UNDER FREE FLOW SITUAITON AT ROAD SEGMENTS BETWEEN GROUPS.

Time frames	**Mean of speeds (km/h)**		**Difference**	**Sig. (ANOVA)**	**Sig. (t-test)**
	Drivers involved in crashes	***Drivers not involved in crashes***			
A.M. peak*	41.914	38.939	2.975	0.003	0.001
Morning	41.820	40.609	1.211	0.168	0.094
Afternoon	42.578	41.659	0.919	0.159	0.054
P.M. peak**	41.300	39.954	1.346	0.094	0.039
Night*	42.801	40.609	2.192	0.001	0.000

* Indicates a significant mean difference (α = 0.05) based on both methods.

** Indicates a significant mean difference (α = 0.05) based on paired-sample t-test.

V. CONCLUSIONS

Many urban traffic surveillance systems have been built in these years. With these systems, data of both objective vehicles and vehicles nearby could be obtained. This current study made use of urban traffic surveillance data, compared the preceding vehicle and following vehicle in order to exclude effects of the external environment, and evaluated the differences between crash-involved and crash-not-involved drivers through headways and speeds.

The conclusions of this paper could be outlined as follows: at most times, drivers who were involved in crashes drive at significantly lower headways through intersections than crash-not-involved drivers, and the differences stably existed several days before the crashes except for crash-involved female drivers which the differences existed occasionally. At morning time and night time, crash-involved drivers' headways at road segments were significantly lower than those of crash-not-involved drivers. At a.m. peak time and night time, crash-involved drivers tended to drive at significantly higher speed than crash-not-involved drivers under free flow situation at road segments.

There are still some shortcomings in this study. In this research, we distinguished different drivers by vehicle license plate. In fact, several drivers might drive the same vehicle at different time. This is quite common in operating vehicle drivers, and it might cause a potential bias to the result.

Identifying at-risk drivers is still a challenging issue nowadays, and this current results suggest that there is a great potential to taking advantage of urban traffic surveillance data to carry out researches.

ACKNOWLEDGMENT

The work described in this paper was supported by grants from the National Natural Science Foundation of China (Grant No. 71301083), the National Basic Research Program of China (973 Project; No.2012CB725405), and the Research Funds of Tsinghua University (No. 20151080412).

We would like to thank Prof. Wong S.C. from the University of Hong Kong for providing suggestions and insights for this research.

REFERENCES

[1] B. E. Sabey and H. Taylor, "The Known Risks We Run: The Highway," in Societal Risk Assessment: How Safe is Safe Enough? R. C. Schwing and W. A. Albers, Eds. Boston, MA: Springer US, 1980, pp. 43-70.

[2] D. Solomon, "Accidents on main rural highways related to speed, driver, and vehicle," 1964.

[3] L. Aarts and I. Van Schagen, "Driving speed and the risk of road crashes: A review," Accident Analysis & Prevention, vol. 38, pp. 215-224, 2006.

[4] A. Baruya, "Speed-accident relationships on European roads," in 9th International Conference on Road Safety in Europe, 1998.

[5] M. C. Taylor, D. A. Lynam and A. Baruya, The effects of drivers' speed on the frequency of road accidents: Transport Research Laboratory Crowthorne, 2000.

[6] T. Brijs, G. Wets, R. Krimpenfort, and C. Offermans, "The impact of hourly measured speed on accident risk in the Netherlands: results from an exploratory study using GIS,", 2006.

[7] C. N. Kloeden, J. McLean and G. F. V. Glonek, Reanalysis of travelling speed and the risk of crash involvement in Adelaide South Australia: Australian Transport Safety Bureau, 2002.

[8] J. Jun, R. Guensler and J. Ogle, "Differences in observed speed patterns between crash-involved and crash-not-involved drivers: Application of in-vehicle monitoring technology," Transportation research part C: emerging technologies, vol. 19, pp. 569-578, 2011.

[9] J. Paefgen, T. Staake and E. Fleisch, "Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data," Transportation Research Part A: Policy and Practice, vol. 61, pp. 27-40, 2014.

[10] H. Lahrmann, N. Agerholm, N. Tradisauskas, T. Næss, J. Juhl, and L. Harms, "Pay as You Speed, ISA with incentives for not speeding: A case of test driver recruitment," Accident Analysis & Prevention, vol. 48, pp. 10-16, 2012.

[11] J. Jun, "Potential Crash Exposure Measures Based on GPS-Observed Driving Behavior Activity Metrics,": Citeseer, 2006.

[12] K. Vogel, "A comparison of headway and time to collision as safety indicators," Accident analysis & prevention, vol. 35, pp. 427-433, 2003.

[13] Transportation Research Board, Highway Capacity Manual, 2010.