On the Effect of Traffic and Road Conditions on the Drivers' Behavior: A Statistical Analysis

Abdalla Abdelrahman
Electrical and Computer Engineering
Queen's University
Ontario, Kingston
Email: a.abdelrahman@queensu.ca

Najah Abu-Ali
College of Information Technology
United Arab Emirates University
Al-Ain, UAE
E-mail: najah@uaeu.ac.ae

Hossam S. Hassanein
School of Computing
Queen's University
Ontario, Kingston
Email: hossam@cs.queensu.ca

***Abstract*—In the last decade, naturalistic driving studies (NDSs) have given researchers an unprecedented way to study the behavior of drivers through the deployment of, and capturing the data from, on-board vehicle sensors and cameras. The ability to determine the dominant driving risk factors can play an essential role in shaping transportation policies and education programs for drivers. This paper presents a cohort study statistical analysis to determine the risks associated with traffic and road surface conditions, quantified in terms of crash and near crash events. Two risk quantification measures, odds ratio (OR) and relative risk (RR), are utilized to signify the associated risk. For this research we used the 100-CAR data set, with a total of 829 crash and near crash and 19616 baseline events, which are driving events captured randomly in normal driving episodes. In the 100-CAR data set, traffic density is divided into six levels according to the traffic flow condition. Similarly, road surface condition is divided into four categories. To quantify the statistical significance of the results, measures such as the *p-value* are employed . The results show that icy roads with level-of-service (LOS) A, wet roads with LOS D, and dry roads with LOS D have the highest risk for crashes and near crashes. These results are proven to be of statistical significance.**

***Index Terms*—Naturalistic driving studies (NDSs), driving behavior, driving risk management, data driven applications, intelligent transportation systems (ITS).**

I. Introduction

According to World Health Organization (WHO), more than 1.25 million individuals lose their lives annually due to road crashes, and between 20 and 50 million people suffer from traffic-related injuries. Moreover, road crashes are expected to remain one of the top ten leading causes of fatalities by the year 2030, if no substantial action is taken [1].

As a result, researchers during the last two decades have been developing different research techniques to study the factors that may affect the crash rate. Among the wide range of research approaches, naturalistic driving studies (NDSs) have recently prevailed [2]–[6]. NDSs give researchers the opportunity to study the behavior of drivers and to explore the driving risk associated with many driving-related elements [7]–[9]. By deploying unobtrusive instrumentation (e.g. radar sensors, OBDII dongles, GPS, forward facing and rear-view video cameras) inside vehicles of recruited volunteer drivers, data about the driver's behavior, the environment, and the vehicle is continuously recorded [2], [3]. The data collected not only gives the chance to study the prevalence of different factors during risky events but also the prevalence of these factors through normal day to day driving episodes, which enables the conduction of statistically sound studies. Results of NDSs are essential tools for transportation policymakers to design safer roads, enact evidence-based driving laws and develop more effective driver education for novice drivers [10]. Moreover, NDSs have also been used to model the behavior of drivers (i.e., tailgating, and lane changing behavior) for predicting and detecting risky events [11] and incorporate these models in self-coaching driving systems [12].

Among the NDSs that were performed, the 100 Car Naturalistic Driving Study (100-CAR NDS) from Virginia Tech Transportation Institute (VTTI) sponsored by the US National Highway Traffic Safety Administration (NHTSA) has become a landmark study with approximately 43,000 hours of recorded data from 241 primary and secondary drivers, using 100 cars [2]. The 100-CAR NDS has a total of 829 crash and near crash events as well as 19616 day to day normal driving recorded events. These recorded events and the associated factors that were present during the events have helped researchers to better understand driver behavior, the causes of crashes, and to develop countermeasures for preventing crashes and near crashes.

Driving risk elements can be divided into three categories. First, driving behavioral factors, for example speeding, sudden braking, distractions, and reckless driving. Second, environmental elements such as road infrastructure, road surface condition, weather, and traffic density. Finally, vehicle-related elements, the vehicles age, and mechanical condition. Although much attention has been dedicated to study the effect of driving behavioral factors on crashes and to quantify the risk associated with each of them, only a few studies were interested in investigating the risk associated with other factors; this is because more than 90% of the crashes are attributed to human error [13]. However, some of these errors can be directly attributed to inconvenient environmental conditions such as traffic density and road surface conditions. The research has shown that the level of driver frustration depends on the level of traffic density [2]. High levels of frustration makes drivers prone to behavioral errors based on how much traffic is present. Similarly, road surface condition can play a significant role in driving behavior errors, depending on the level of

978-1-5386-2070-0/18/$31.00 ©2018 IEEE

experience a driver has in dealing with such conditions.

This paper presents a retrospective cohort-study statistical approach to investigate the risk associated with different traffic density levels and road surface condition categories. Generally speaking, retrospective cohort studies represent a research approach in which causes of an experiment outcome are investigated and links between risk factors and outcomes are established, for already gathered data.

Odds ratio (OR) and relative risk (RR) are used to examine the associated risk of each factor. Both crash and near crash events are utilized as risk indicators because of the limited number of crashes in the 100-CAR NDS. The remainder of this paper is organized as follows. In section II, some of the related work presented in the literature is shown. Section III explains the statistical methodology adopted. In Section IV, the results are presented and discussed. Finally, conclusions are drawn in section V.

II. Related Work

Studying the factors that influence the crash risk is a multidisciplinary field of research that converges at the intersection of behavioral psychology, statistics, transportation engineering, and data science. Throughout the last decade, many notable studies have focused on examining the different elements and driving attributes that may lead to accidents. Authors in [14] collected naturalistic driving data to study the triggering factors that indicate the future occurrence of risky events. They utilized near crash events as a substitute measure for riskiness rather than crashes. The work presented in [15] utilizes the Crash Record Information System (CRIS) database from the Texas Department of Transportation (TxDOT) to determine the main features of crashes that involve pedestrians. A classification and regression tree (CRT) analysis was performed to figure out which factors had the most influence affecting the severity of these crashes. It was concluded that lighting conditions, road class, traffic control, right shoulder width, the involvement of a commercial vehicle, pedestrian age, and the collision manner are the most influencing factors. Authors in [16] utilized the 100-CAR NDS to study the effects of driver distraction on the probability of crashes/near crashes. Driver distraction has been quantified solely on the cumulative off-road glance duration measured by a camera focused on the drivers eyes, and it was found that it is linearly proportional to the risk of crash and near-crash events. Moreover, it was shown that traffic density is a significant moderator to this relationship. Hasan et al. in [17] adopted K-means clustering algorithm to group near crash events according to their driving risk. Three variables, deceleration, braking pressure, and headway time, have been used as risk indicators in these events. Then, an ordered logit regression model has been utilized to study the main contributing factors that affect the driving risk of near crash events. The study was conducted from an NDS collected in Wuhan city in China with a total of 1670 near crash events. The results indicated that road condition, time of day, the day of the week, age and driving experience are significant in risk determination. However, traffic density has been put under only two categories, congested and non-congested and most of the results presented possess marginal statistical significance.

Despite the research efforts mentioned above, to the best of our knowledge, no study has solely investigated the risk associated with traffic density or jointly when combined with different categories of road condition, and with this magnitude of data.

III. Methodology

A retrospective cohort study approach is followed to determine the association between risk factors (i.e., traffic and surface conditions) and the occurrence of crash and near crash events. A cohort, rather than case-control approach, is utilized since the risk factors have already been pre-assumed.

In this study, two groups of events are defined. The exposure group represents the group of events either normal or safety-critical in which drivers are exposed to the risk factors of interest. The second group is the control group which contains the set of events whether normal or safety-critical in which drivers are not exposed to the risk factors of interest.

A total of 19616 baseline driving events, which are driving events captured randomly in normal driving episodes, are used to reflect the exposure rate of the risk factors during safe driving events. Meanwhile, a total of 829 safety-critical events are employed to determine the exposure rate of these factors during crash and near crash events. The extraction of safety-critical events was performed by Virginia Tech Transportation Institute (VTTI) after the conclusion of the 100-CAR NDS project.

Analyses are performed to calculate the risk associated with six different traffic density levels (A-F) as well as four road surface conditions categories. The detailed operational definition for each of these categories is listed in tables I and II.

TABLE I
Definition of traffic density levels

Traffic condition	**Definition**
Level-of-service A	Drivers are free to pick the desired speed and to maneuver.
Level-of-service B	Relatively small decline in the freedom of speed and maneuverability compared to LOS A.
Level-of-service C	The decline in the level of moving comfort is noticeable. However, it is the zone of stable traffic flow.
Level-of-service D	Traffic is dense but stable. Maneuverability and speed comfort level severely declines.
Level-of-service E	Unstable flow with minimal and uniform speeds. Drivers tend to be very frustrated.
Level-of-service F	Breakdown flow with stop-and-go cyclic waves.

TABLE II
DEFINITION OF ROAD SURFACE CONDITION CATEGORIES

Road condition	Definition
Dry	No foreign material (snow, ice, oil, water) on the road in the area of the event.
Wet	Road is partially or completely wet in the area of the event.
Snowy	Snow or slush on the road in the area of the event.
Icy	Un-melted ice on the road in the area of the event.

A. Risk association measures

OR and RR risk association measures have been adopted in this work. Although OR asymptotically approaches RR when the unexposed incidences are much larger than the exposed ones, they both have different statistical interpretations. The OR in the context of this paper refers to the ratio of the odds that a crash/ near crash event will happen given the exposure to a certain traffic and/or road surface condition, to the odds of a crash or near crash event happening in the absence of these exposures. To calculate the odds ratio, four values are determined as shown in table III. The first value, "a", is the number of crashes and near crashes in which the subject driver (SV) is exposed to the risk factor. The value "b" refers to the normal driving events in which the SV driver is also exposed to the same risk factor. Conversely the values "c" and "d" represent, respectively, the number of crashes and near crashes and normal driving episodes in which the SV driver is not exposed to the risk factor of interest.

TABLE III
DEFINITION OF OR & RR VARIABLES

	Exposed	Not exposed
Crash or near crash event happens	a	c
Crash or near crash event does not happen	b	d

The OR is then calculated using the following equation:

$$OR = \frac{(a/c)}{(b/d)} \tag{1}$$

The value of the OR determines how much the exposure to the risk factor affects the odds of the crashes and near crashes. If the OR possesses a value of 1, the exposure to the risk factor does not influence the odds of the crashes and near crashes. On the other hand, an OR of a value greater than 1 means that the exposure to the risk factor is associated with higher odds of crashes and near crashes. Finally, an OR with a value less than 1 mean reflects lower odds of the crashes and near crashes occurring given the exposure to the risk factor.

To estimate the precision of the OR, its 95% confidence interval (CI) is calculated. Upper and lower 95% CI limits are found in terms of the standard error (SE) of the log odds ratio (LOR). The following equation was adopted to calculate the SE:

$$SE[\ln(OR)] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \tag{2}$$

The upper and lower 95% CI limits are then calculated according to the following equations:

$$95\%\ CI_{Upper} = \exp^{\ln(OR)+1.96\times SE(\ln(OR))} \tag{3}$$

$$95\%\ CI_{Lower} = \exp^{\ln(OR)-1.96\times SE(\ln(OR))} \tag{4}$$

Several statistical significance tests can be utilized for the OR. The most common is the Fisher's exact ratio which gives an exact calculation for the *p-value*. It is expressed mathematically as:

$$p = \frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{n}{a+b}} \tag{5}$$

where $n = a + b + c + d$. In this paper, an approximation has been adopted to calculate the *p-value* since the sample size is relatively large. Results are considered statistically significant when the *p-value* possesses a value less than 0.05, marginally significant between 0.05 and 0.1, and not significant otherwise.

Another widely used risk association measure is the RR. When the number of cases (i.e., crashes and near crashes) is large, RR is proven to give more accurate results. RR is the ratio of two probabilities. The first is the probability of having a crash/ near crash event given the exposure, while the latter is the probability of a crash or a near crash event given the absence of the exposure. If a crash or a near crash event is denoted by C, and the exposure to the risk factor is denoted by E, then RR can be expressed mathematically as:

$$RR = \frac{Pr(C|E)}{Pr(C|E')} = \frac{Pr(C \cap E) \times Pr(E')}{Pr(C \cap E') \times Pr(E)} \tag{6}$$

which can simply be written as:

$$RR = \frac{a/(a+c)}{b/(b+d)} \tag{7}$$

IV. RESULTS AND DISCUSSION

Driver frustration and driving behavioral errors can be directly attributed to certain traffic and road surface conditions. In this section, three sets of results are obtained to indicate the risk of being exposed to different levels and categories of traffic and road conditions. For this purpose, both crash and near crash events are used as risk reflectors. All crash and near crash events of the 100-CAR-NDS dataset are included in this study, even those in which the SV driver is at fault. This is because the goal of this research is to measure the associated risk regardless of who was at fault or responsible for the crash or near crash.

The first set of results presents the association between the exposure to six unique traffic density levels and the crash and near crash rate. The second set of results shows the influence of being exposed to four different road surface condition categories on the crash/ near crash rate. The last set studies the joint influence of the exposure to both traffic density levels and road surface conditions simultaneously on having a crash/ near crash event.

A. *Traffic density*

Figure 1 depicts the traffic density forest plot. LOS A, which corresponds to free traffic flow, is associated with a low driving risk with an OR of 0.37 (< 1) and 95% CI ranging from 0.31 to 0.43. Similarly, LOS A RR possesses a value of only 0.54 which confirms its low risk association as shown in table IV. In this case, the very small *p-value* (< 0.0001) shows that these results are statistically significant. Table IV shows values for the different exposure and control groups variables for each traffic density LOS.

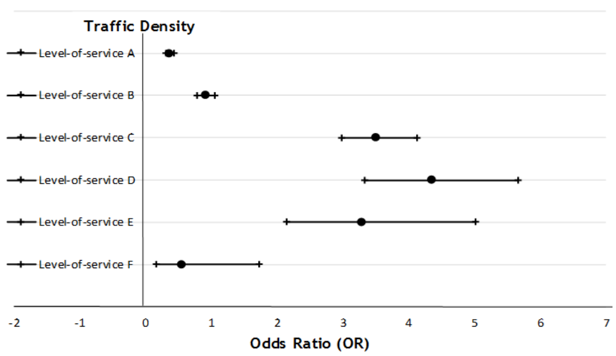

Fig. 1. Forest plot of traffic density levels ORs

LOS B possesses an OR value of 0.92. Although higher than the LOS A OR value, the OR value is still less than 1 which means it is also associated with low odds of the crashes and near crashes. The OR 95% CI shows that it spans 1, which means that this result does not reach statistical significance as confirmed by the high *p-value* (0.2515). The RR again coincides with the OR value with a value < 1.

In contrast, the exposure to LOS C, LOS D, and LOS E seems to elevate the driving risk with OR values of 3.5, 4.4 and 3.3, and RR values of 2.86, 4.07, and 3.22, respectively. All of these results are of statistical significance since they possess *p-values* < 0.05. Finally, LOS F possesses OR and RR values of 0.55. These results are statistically insignificant since the *p-value* in this case is much greater than the marginal value of 0.1.

B. *Road surface conditions*

Table V shows event numbers in the exposed and control groups for each surface condition category. The table shows

TABLE IV
TRAFFIC DENSITY STATISTICS

Traffic Condition	**a**	**b**	**c**	**d**	***p-value***	**RR**
Level of service A	220	9705	607	9911	<0.0001	0.54
Level of service B	297	7432	530	12184	0.2515	0.95
Level of service C	213	1764	614	17852	<0.0001	2.86
Level of service D	69	402	758	19214	<0.0001	4.07
Level of service E	25	184	802	19432	<0.0001	3.22
Level of service F	3	129	824	19487	0.3069	0.55

the relatively high number of crashes and near crashes that occurred when the road surface was dry (i.e., 714). The road surface condition is a contributing factor in almost 86% of the total number of crashes and near crashes. However, this may be attributed to the large number of baseline events in which the SV driver was exposed to (i.e., 17582 events on dry roads). Yet, only 7 crash and near crash events took place when the road was icy. Nevertheless, drivers were reported to be exposed to icy roads in only 0.0005% of the total baseline events (i.e., 10 out of 19616).

TABLE V
SURFACE CONDITION CATEGORIES VARIABLES

Surface Condition	**a**	**b**	**c**	**d**
Dry	714	17582	114	2034
Icy	7	10	821	19606
Snowy	4	190	824	19426
Wet	102	1828	726	17788

As shown in table VI, although 86% of the crashes/ near crashes occurred when the road was dry, the OR value is only 0.725 and the RR $\simeq 1$. This means that crashes and near crashes are more likely to occur in the control group rather than in the exposed group where drivers are exposed to dry roads. This result is considered statistically significant as the *p-value* in this case is less than 0.05. On the contrary, the exposure to icy and wet roads seems to be associated to high odds of risky events.

TABLE VI
ROAD SURFACE CONDITIONS STATISTICS

Surface Condition	**OR**	**CI Upper (95%)**	**CI Lower (95%)**	***p-Value***	**RR**
Dry	0.725	0.8875913	0.5914786	0.0019	0.962
Icy	16.72	44.027381	6.3469475	<0.0001	16.58
Snowy	0.496	1.3393445	0.1839216	0.1666	0.499
Wet	1.37	1.6912973	1.1051189	0.0040	1.322

TABLE VII
JOINT RISK ASSOCIATION OF TRAFFIC DENSITY AND ROAD SURFACE CONDITION

Road and Surface Conditions	**a**	**b**	**c**	**d**	**OR**	**CI Upper (95%)**	**CI Lower (95%)**	***p-Value***	**Z-Statistic**	**RR**
Level-of-service A & Dry	182	8711	646	10906	0.352725087	0.41678281	0.298512761	<0.0001	12.239	0.495000491
Level-of-service A & Icy	5	6	823	19611	19.85722965	65.19966549	6.047723809	<0.0001	4.927	19.74335749
Level-of-service A & Snowy	4	138	824	19479	0.685204728	1.856150886	0.252945772	0.4572	0.744	0.686725478
Level-of-service A & Wet	29	845	799	18772	0.806315587	1.175373319	0.553139003	0.2629	1.12	0.81309922
Level-of-service B & Dry	264	6668	564	12949	0.909003306	1.055178252	0.783078128	0.2098	1.254	0.938016744
Level-of-service B & Icy	1	4	827	19613	5.928960097	53.10652317	0.661925612	0.1116	1.591	5.923007246
Level-of-service B & Wet	32	722	796	18895	1.052074778	1.509644057	0.733193586	0.7829	0.276	1.050062227
Level-of-service C & Dry	186	1583	642	18034	3.300570902	3.916513175	2.781496651	<0.0001	13.678	2.783775989
Level-of-service C & Wet	26	172	802	19445	3.665037986	5.569247353	2.411906419	<0.0001	6.084	3.581353219
Level-of-service D & Dry	58	352	770	19265	4.12252804	5.492646279	3.094180214	<0.0001	9.675	3.903800231
Level-of-service D & Wet	11	49	817	19568	5.37676417	10.37875045	2.785459874	<0.0001	5.013	5.318618752
Level-of-service E & Dry	22	135	806	19482	3.939012958	6.215473524	2.496321966	<0.0001	5.891	3.860923242
Level-of-service E & Wet	3	27	825	19590	2.638383838	8.71474932	0.798768734	0.1115	1.591	2.632447665
Level-of-service F & Dry	2	115	826	19502	0.410611643	1.664549214	0.101289839	0.2126	1.246	0.412035287
Level-of-service F & Wet	1	13	827	19604	1.823458283	13.95604437	0.238248032	0.5629	0.579	1.822463768

Table VI shows that the risk of being exposed to icy roads is higher by 16.52 times when compared to other road surface categories. Similarly, the odds of having a crash or a near crash event when driving on wet roads is approximately 1.37 times higher than the risk imposed when driving on other roads. Similar to dry roads case, these results are also of statistical significance since the *p-values* < 0.05.

Surprisingly, there was no association between the risk of having a crash/ near crash and driving in snowy roads since the OR and the RR values are less than 1. Perhaps this could be attributed to the cautious attitude of drivers when they drive in snowy environments. The OR in this case does not reach statistical significance, however, it can be said to possess a marginal statistical significance with a *p-value* of 0.16.

C. Traffic density and road surface conditions

In this work, The joint risk association of both traffic density and road surface conditions has also been studied. The goal is to examine the crash or near crash risk for different traffic density levels and road surface categories, happening simultaneously.

Table VII depicts the results for 15 joint cases. Although the results were calculated for all 24 cases, the remaining 9 are not presented here since there were not enough events in either one or more variable fields. As shown in the table, the case where drivers are exposed to icy roads with LOS A possesses the highest OR and RR values with a *p-value* < 0.0001. To put this finding into perspective, although captured in only 6 baseline events, the exposure to this type of traffic and surface condition has contributed in 5 risky events. That explains the high OR and RR values of 19.86 and 19.74, respectively.

Similar cases that possess high risk associations and are statistically significant are the cases of wet roads & LOS D, dry roads & LOS E, wet roads & LOS D, dry roads & LOS D, wet roads & LOS C and dry roads & LOS C. Among these cases, wet roads & LOS D reaches the highest OR and RR values of 5.38 and 5.31, respectively. Two cases are also found to be associated to driving risk, however, they are marginally significant. These are the cases of icy roads & LOS B and wet roads & LOS E, where the first possesses the highest OR and RR values. The discrepancy between OR and RR values in some cases (e.g. dry roads & LOC C) is due to the relatively large number of exposed cases (large RR probability values). However, they both have values greater than 1 (3.3 and 2.7, respectively). Six of the nineteen reported cases do not reach statistical significance because their 95% CI spans the neutral value of 1 and their *p-value* > 0.1. These cases are: snowy road & LOS A, wet road & LOS A, dry road & LOS B, wet road & LOS B, dry road & LOS F and wet road & LOS F.

V. Conclusion

This paper presented a cohort-study statistical analysis to determine the risk associated with different traffic density levels and road surface condition categories. Both crash and near crash events were adopted as risk indicators. For our research, data-sets from the 100-CAR NDS project were utilized. The event video reduced dataset contains the information regarding 829 crash and near crash events, whereas the baseline video reduced data set comprises the information for 19616 baseline events. Baseline events were used to indicate the prevalence of different traffic density and road condition levels during normal driving episodes.

With respect to traffic density, the results show that driving in stable traffic flow but with limited control of speed and maneuverability (LOS D) poses the highest OR and RR values. Considering road condition, the results indicate that driving on icy roads results in the highest associated risk. Contrary to popular belief, driving in snowy environments was identified as a low driving risk. The joint analysis when considering the simultaneous effect of both traffic and different road surface condition levels show that driving on an icy road with stable flow (LOS A) has the highest OR and RR values. The analyses show that all of the aforementioned results are either statistically significant or have marginal statistical significance. Cases such as driving on icy roads with traffic LOS D were disregarded because of the limited number of occurrences, either during baseline, or crash and near crash events. In the future, using larger scale dataset, such as the Strategic Highway Research Program 2 (SHRP2) may provide more insights on the risk associated with such cases.

References

[1] World Health Organization (WHO) "Global status report on road safety 2015 ",World Health Organization, 2015

[2] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, and J. Jermeland, "The 100-car naturalistic driving study, Phase II - results of the 100-car field experiment ", Washington, DC:NHTSA.

[3] T. A. Dingus, J. M. Hankey, J. F. Antin, S. E. Lee, L. Eichelberger, K. E. Stulce, D. McGraw, M. Perez, L. Stowe, "Naturalistic driving study: Technical coordination and quality control ", *SHRP2 Rep. S2-S06-RW-1*, Transportation Research Board, Mar. 2015.

[4] M. Benmimoun, A. Ptz, A. Zlocki, and L. Eckstein, "euroFOT: Field operational test and impact assessment of advanced driver-assistance systems: Final results", in *Proc. FISITA World Automotive Congr. Lecture Notes in Electrical Engineering*, vol. 197, pp. 537547, 2013.

[5] R. Eenink, Y. Barnard, M. Baumann, X. Augros, and F. Utesch, "UDRIVE: The European naturalistic driving study", presented at the Transport Research Arena Conf., Apr. 2014.

[6] M. A. Regan, A. M. Williamson, R. Grzebieta, J. L. Charlton, M. G. Lenn, B. Watson, N. L. Haworth, A. Rakotonirainy, J. E. Woolley, R. W. G. Anderson, T. M. Senserrick, and K. L. Young,"Australian 400-car Naturalistic Driving Study: Innovation in road safety research and policy", in *Proc. Australian Road Safety Research, Policing & Education Conf.*, 2013, pp. 113.

[7] C. Carney, D. McGehee, K. Harland, M. Weiss, M. Raby,"Using Naturalistic Driving Data to Assess the Prevalence of Environmental Factors and Driver Behaviors in Teen Driver Crashes ", *AAA Foundation for Traffic Safety*, Washington, DC, 2015.

[8] L. Precht, A. Keinath, J.F. Krems,"Identifying the main factors contributing to driving errors and traffic violations Results from naturalistic driving data", *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 49, pp. 49-92, Aug. 2017

[9] L. Precht, A. Keinath, J.F. Krems,"Identifying effects of driving and secondary task demands, passenger presence, and driver characteristics on driving errors and traffic violations Using naturalistic driving data segments preceding both safety critical events and matched baselines", *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 51, pp. 103-144, Nov. 2017.

[10] T. A. Dingus,F. Guo, J., S. Lee, J. F. Antin, M. Perez, M. Buchanan-King, J. Hankey, "Driver crash risk factors and prevalence evaluation using naturalistic driving data ", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. 2636-2641, Mar. 2016.

[11] C. Miyajima; K. Takeda, "Driver-Behavior Modeling Using On-Road Driving Data: A new application for behavior signal processing ", *IEEE Signal Processing Magazine*, vol. 33, issue 6 pp. 14-21, Nov. 2016.

[12] K. Takeda, C. Miyajima, T. Suzuki, P. Angkititrakul, K. Kurumida, Y. Kuroyanagi, H. Ishikawa, R. Terashima, T. Wakita, M. Oikawa, Y. Komada, "Self-Coaching System Based on Recorded Driving Data: Learning From Ones Experiences ", *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, issue 4, pp. 1821 - 1831, Dec. 2012.

[13] T. G. Brown, M. C. Ouimet, M. Eldeb, J. Tremblay, E. Vingilis, L. Nadeau, J. Pruessner, A. Bechara, "The effect of age on the personality and cognitive characteristics of three distinct risky driving offender groups ", *Personality and Individual Differences*, vol. 113, pp. 48 - 56, July 2017.

[14] Y. Zheng, J. Wang, X. Li, C. Yu, K. Kodaka, k. Li, "Driving risk assessment using cluster analysis based on naturalistic driving data ", *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 2584-2589, Nov. 2014.

[15] V. Iragavarapu, D. Lord, K. Fitzpatrick, "Analysis of Injury Severity in Pedestrian Crashes Using Classification Regression Trees ",*Transportation Research Board 94th Annual Meeting*, Jan. 2015.

[16] R. Tian, L. Li, M. Chen, Y. Chen, G. J. Witt, "Studying the Effects of Driver Distraction and Traffic Density on the Probability of Crash and Near-Crash Events in Naturalistic Driving Environment ", *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, issue 3, pp. 1547-1555, Sep. 2013.

[17] R. Tian, L. Li, M. Chen, Y. Chen, G. J. Witt, "Examining contributing factors on driving risk of naturalistic driving using K-means clustering and ordered logit regression ", *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, Banff, Canada, pp. 1189-1195, Aug. 2017.