Relationship between Socio-Demographic of Drivers and Traffic Violations and Crashes Involvements

Mohamed Shawky[1, 2], Yousef Al-Badi[1], Abudullah Al-Ghafli[1]
[1]Abu Dhabi Traffic Police
Abu Dhabi, UAE
[2]Ain Shams University
Cairo, Egypt
m_shawky132@hotmail.com; abdulla522@gmail.com

Abstract - This paper aims to investigate the interrelationships between the socio-demographic characteristics of the drivers and their behaviour in terms of the historical records of traffic rule violations and crashes. Individual driver's data were collected and integrated from four different databases over eight years from 2008 to 2015. A total of 608,611 individual driver's records of violations and crashes were employed in the analysis. The examined drivers' demographics variables included: gender, age, nationality, number years of driving experience, occupation, level of education, marital status and number of vehicles owned by the driver. Drivers have been grouped based on their demographics and violation records and then the crash rates of both property damage only (PDO) crashes and severe crashes were calculated for each group. The analysis showed a strong relationship between number of drive's violations and crash rates. In addition, the demographic characteristics of the drives have a significant impact on the value of the crash rates. male, young (18-24 years) older drivers (greater than 65 years), local nationality, high number of owned vehicle by driver, low educated and drivers who work as a driver have higher crash rates compared to others.
Two different models were developed to estimate the variables that significant affect the frequency occurrence of the PDO and severe crash of drivers by applying the Negative Binomial Regression approach. The results confirmed the findings of the two-way data analysis regarding the drivers' demographics. In addition, it showed that some variables are significant in the occurrence of the PDO crashes and not significant in severe crashes such drivers' nationality (Asians), mobile usage and seat-belt violators. in addition, the results showed that speeding and tailgating violation types significantly affect the occurrence of the both types of crashes.

Keywords: Socio-demographics of drivers, Crash Rate, Traffic violations, Risky driving behaviour, Crash probability, Negative binomial model, Road safety

1. Introduction

Most the crash reports stated that the human errors are considered as the main or secondary cause of the traffic crashes. Some ninety percent of motor vehicle crashes are caused at least in part by human error. however, the percentage share of human errors in the occurrence of traffic crashes is vary among cities based on the socio-demographics characterises and driver's behaviour. Treat et al. found that "human errors and deficiencies" were a definite or probable cause in 90-93% of the examined crashes [1]. More updated report by NHTSA's 2008 stated that human error is the critical reason for 93% of crashes [2]. It has also been estimated that human actions are a sole or contributory factors in as much as 95% of traffic crashes [3; 4]. Despite these estimates, it is apparent that, in contrast to other domains in which human behaviour has been identified as a major problem, the interaction between drivers' characteristics and crash involvements has previously received only limit attention from road safety researchers.

The main objective of the current paper is to investigate the interaction between the driver socio-demographics characteristics in terms of gender, age, nationality, education level, occupation (driver or other), number of vehicles owned for the driver and marital status with the drivers behaviour in terms of their historical records of total traffic rule violations and records of some types of traffic violations such as speeding, seat belt usage, mobile usage and also their records of both types of crash involvements (i.e., severe crashes and property damage only crashes). Historical data over eight years of individual drivers were utilized in this paper.

2. Literature Review

Researchers have divided the human factors into three different groups: error, lapses and violations [5]. Violations are different from errors and lapses, since they are deliberate and can be understood in terms of social and motivational factors such as the persons; attitude and norms. In case of the drivers and traffic crashes, early on, Peck et al. [6] stated that it is difficult to accurately predict drivers who will and will not be involved in crashes. However, one year after, Stewart and Campdell investigated the probability of future crashes in terms of historical crash and/or violation records [7]. This study observed a four-year history of crash and traffic violation records of North Carolina drivers to predict the future crashes. In 1990s' and more recent, several studies had been published to suggest that the involvement in a traffic crash is not a random event and it can be predicted in terms of a combination factors of drivers' socio-demographics characteristics and their historical records of traffic violations and crashes factors.

Two studies were conducted in California that aimed to assess the accuracy of predicting future accidents risk using various combinations of demographic and prior driving records variables [8; 9]. Approximately 140,000 records of licensed drivers including their age, gender and driving record variables were analyzed. About 17 regression models were developed using various combinations of demographics and prior driving variables. The results indicated that the models that use prior total accidents as a predictor variable of future accidents perform better than other models that do not use total accident as a predictor and the same results was found for using historical violation. It also found that adding demographic variables of the drivers to the model increase the performance of the model. Accordingly, the total historical accidents and violation records can be used as a measure to predict high-risk drivers in order to intervene before this risk is realized, which support the current point count strategy in California which attempts to optimize the identification of driver having a high probability of subsequent accidents involvement

Alver et al. [10], explored the interaction between socio-demographic characteristics of traffic rule violators (four types of traffic violations records were aggregated) and accident history for young drivers (18-29 years old) by applying binary logit models. The analysis showed that 23.9% of drivers were involved in at least one traffic accident in last three years. This accident rate increases to 38.3% for those who received at least one traffic violation in last three years and peaks to 47.4% for those who were fined for seat belt violations. In addition, through examining drivers' age impact, previous violation records and crash data, Daigneault et al. [11] concluded that prior crashes would be a better predictor for crash risk than prior violations. It was also found that this relationship increase with elder drivers (drivers have more than or equal 65 years old). Furthermore, Gerber and Peck [12] showed a significant effect of age and prior violations on subsequent of crashes, especially for elder groups of drivers (aged 60-69 and above 70).

Nishida [13] analyzed crash and violation data of 81 million licensed drivers in Japan (approximately 1 million crash record and 10 million violation records per year). It was found that drivers who experience accidents drive more carefully immediately after an accident, revealed high accident rates among drivers who have experienced certain violations, and produced other findings that could constitute a foundation for developing individual driver-targeted measures. In addition, the drivers with a history of numerous accidents or apprehensions/violations are more likely to cause accidents. Chandraratna et al. [14] investigated Kentucky drivers' characteristics to develop a crash prediction model that can be used to estimate the likelihood of a driver being at fault for a near future crash occurrence by using multiple logistic regression technique. The developed model results indicated that the total number of previous at-fault crash involvements, and having previous driver license suspensions and traffic school referrals are strongly associated with a driver being responsible for a subsequent crash. In addition, a driver's likelihood to be at fault in a crash is higher for very young or very old, males, drivers with both speeding and non-speeding violations, and drivers that had a recent crash involvement.

Some researched used the self-reported technique to investigate the interaction between drivers' characteristics and the traffic violations and crashes involvements [e.g., 15 - 17]. For example, Machado-Leon et al. [18] investigated crash risk perceptions in an inter-city, two-way road context of 492 drivers by using a Stated Preference ranking survey. The study that all risky driving behaviour showed a significant potential effect on crash risk perceptions, and model's results allowed to differentiate more important from less important unsafe driving behaviours based on their weight on perceived crash risk. Furthermore, the study analyzed the potential differences in risk perception of these traffic violations between drivers of different characteristics, such as driving experience, household size, income and gender. McDonald et al. [19] investigated behaviour of some groups of drivers and found that that speeding was the most reported act committed in 90% of the cases, young males more than young females. The youngest group was three to five times more likely to be involved in "own fault" road crash than the oldest age group. Some researchers explained the differences of risky driving between men and women due to different driving skills and experiences [20; 21].

3. Data Collection and Preparation

The employed data in this paper was extracted from the Emirate of Abu-Dhabi (AD), the capital of UAE, traffic police databases during eight years from 2008 to 2015. The drivers' data were integrated from different sets of data; 1) traffic violations, 2) property damage only (PDO) crashes, 3) Severe crashes and 4) driver and vehicles license information. Both the PO) crashers and severe crashes (i.e., any crash with at least one injury of fatality) were separately analyzed. Any driver has at least one record of traffic violation and/or crash in the databases during the study period was involved in the analysis process. About 1.1 million licensed drivers are listed in the database. In order to refine the accuracy of the data, a filtration process of data was conducted. The excluded drivers in the used data include three groups: drivers with uncomplete demographics information, drivers with zero records of both violations and crashes, drivers who issued their driving licenses after year 2008. This proses assure that the selected drivers have driving experience over all the eight years of study period. Based on that, the selected data sample include 608,611 individual drivers with full records of demographics, violations and crashes. These drivers have about 16,010 severe crashes, 540,218 PDO crashes and 4,419,404 traffic violations.

It is worth mentioning that the licensed drivers' community in AD has a unique composition where more than two hundred different nationalities. About 85% of the licensed drivers in AD are males, 92% are younger than 45 years old, 13% are local drivers and 87% are foreigners. Such a unique composition gives a chance to analyzed more road safety variables that were not be able to be investigated in prior studies such as the different nationalities with their culture background of drivers who are driving in the same community.

3. Relationships between drivers' demographics and Crash involvements

3.1. Crash rate calculation

For each examined variable of drivers' demographics, the total number of individuals and total number of their severe and PDO crashes are extracted from the database. Each variable was considered as a group of drivers. For example, male drivers are considered as a group "i" after that the total number of male drivers and their crashes and violations records were defined. Then the crash rate of a drive group "i" was calculated as follows:

$$R_i = \frac{y_i}{N_i} \tag{1}$$

Where: i = index for the examined variable (i.e., gender, age, nationality, violation, .….etc.),
R_i = The crash rate per drivers for each group type i,
y_i = the total number of observed severe or PDO crashes for the group of drivers i,
N_i = the total number of drivers belonging to the group i,

3.2. Violation and crash rate relationship

Figures 1-a) and 1-b) show the calculated crash rates of PDO crashes per driver and severe crashers per 1,000 drivers in terms of number of violation of individual driver, respectively. It shows that both the PDO and severe crash rates of divers increases with increasing the number of committed violation of the driver during the eight years of the study period. However, in case of the PDO crashes the trend increases drastically up to about 32 violations (i.e., around 4 violations per year) but after that the crash rate become stable and not increases. For severe crashes the trend increases with strong correlation ($R^2 = 0.97$) up to about 25 violations (i.e., around 3 violations per year) after that the rate of increasing become lower and the correlation become lower ($R^2 = 0.11$). In general, these figures reflect the strong relationship between the driver's historical records of traffic violations and his/her involvements in either PDO or severe crashes. That means the violation records can be used as an indicator for probability of that the driver may involve in a crash in the future. In addition, drivers who have more than 4 violations per year should attend special programs as preventative measurement to enhance their behaviour.
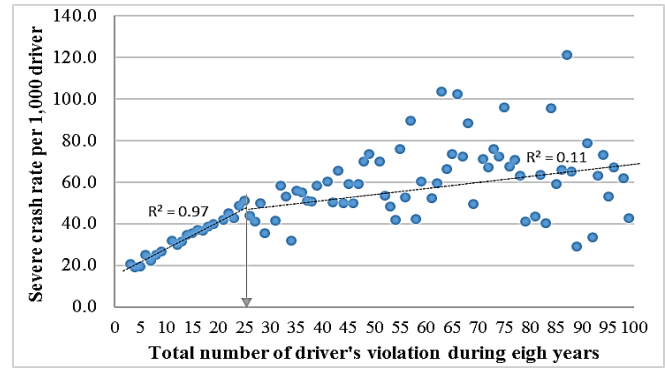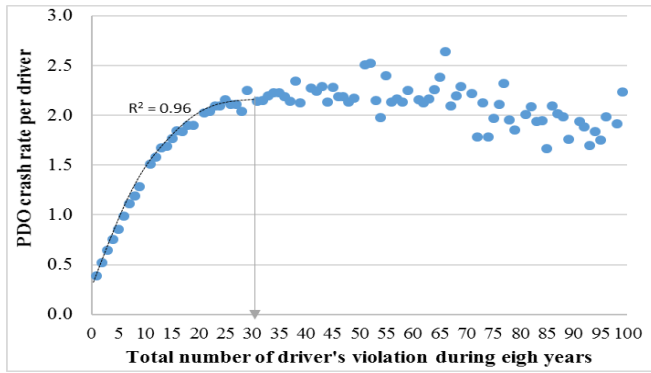
a) PDO crashes rate b) Severe crashes rate

Fig. 1: Crash rate in terms of number of driver's violations.

3.3. Drivers' demographics and crash rate interrelationships

The rate of both PDO and severe crashes for different groups of drivers based on their demographics information were examined. Figure 2 shows the crash rate based on the driver's gender. It shows that male drivers have higher PDO crash rates than female. However, there is no significant difference between male and female divers with respect to the severe crash rates. Figure 3 shows the crash rate in terms of the driver's age group. It shows that the crash rate of young drivers (18-24 years old) and older drivers (greater than 65 years old) have higher crash rate compared to mid-age groups. This finding is consistence with the findings of prior studies that have been discussed in the literature review. The same trend of crash rate is shown in terms of the drivers' year of experiences as shown in Figure 4. That can be explained as the higher year of experience means older drivers.
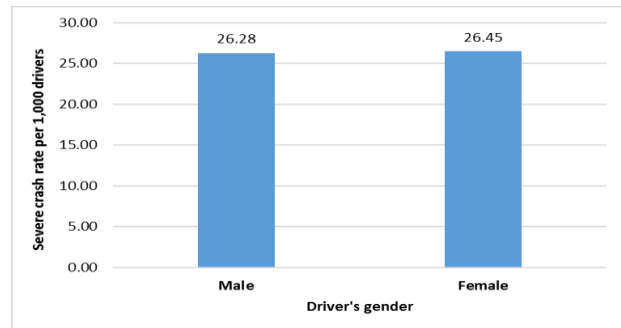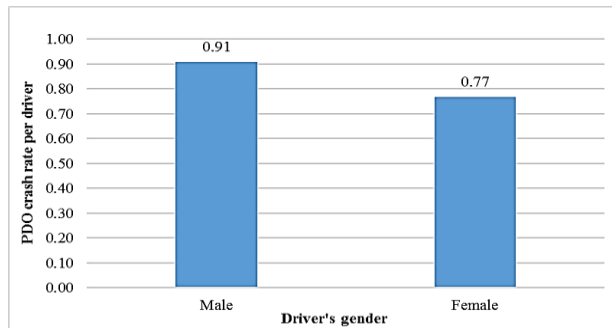

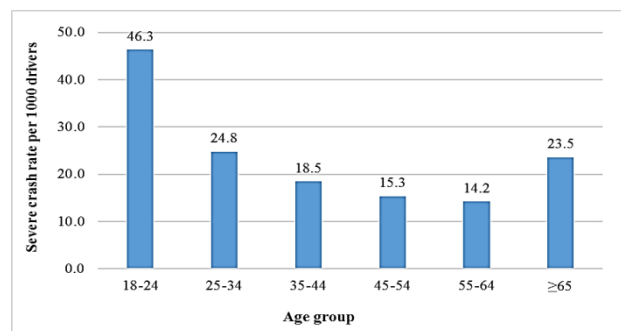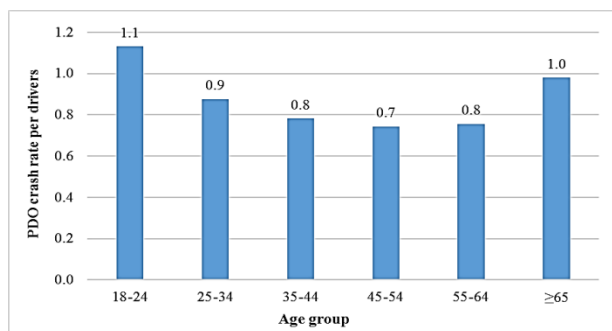Fig. 2: Crash rate in terms of driver's gender.

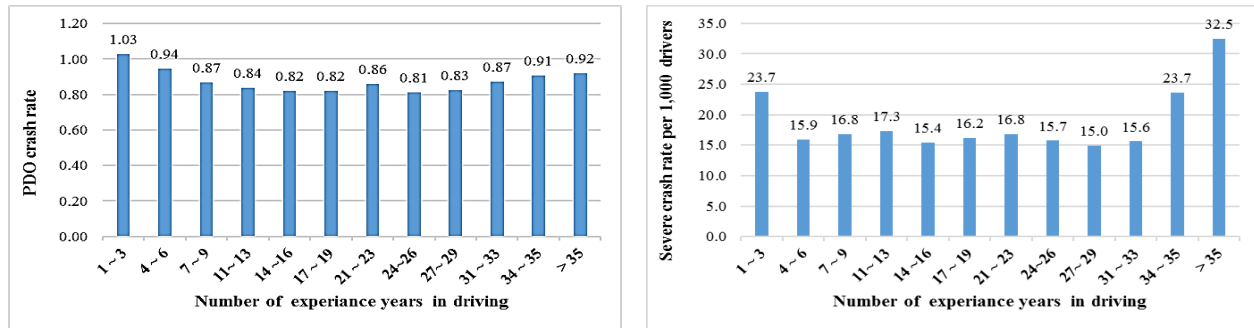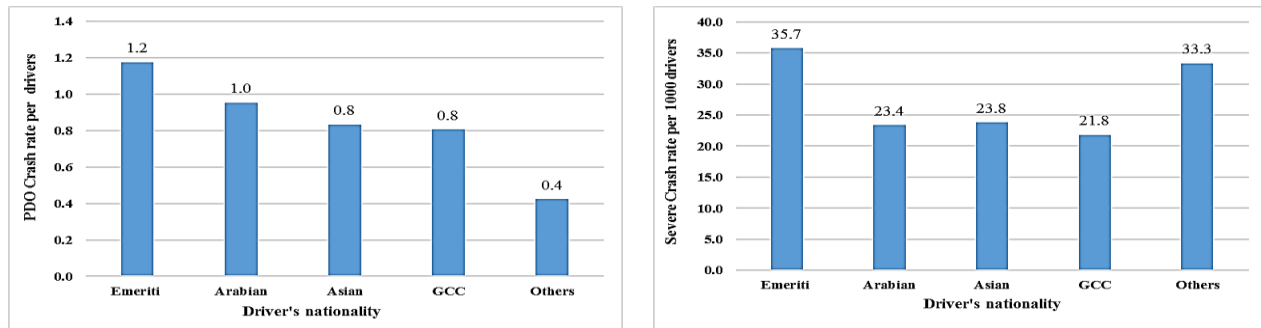Fig. 3: Crash rate in terms of driver's age.

Fig. 4: Crash rate in terms of driver's experience years in driving.

With regard to the drivers' nationalities, Figure 5 indicates that the crash rate is different form nationality to another. It shows that local drivers (i.e., Emeriti) has the highest PDO crash rate compared to other nationalities. That can be explained due to high income of local drivers or due to high travel mileages since most of local people rely on private vehicles in travelling between cities and a significant percentage of locals are living in a city and their work in another city. It also shows that Asians' nationalities have moderate crash rate compared to other nationalities despite they represent the majority of licensed drivers (47% of total drivers). That can be explained due to their law income and high values of traffic fines in AD that push them to be very careful during driving. However, more investigation regarding the driving behaviour among different nationalities are required in order to interpret these findings. Regarding the education level of drivers, Figure 6 shows that high educated drivers (college and graduated degree) have lower crash rates compared to low educated drivers.

* GCC is the Gulf Cooperation Council nationalities

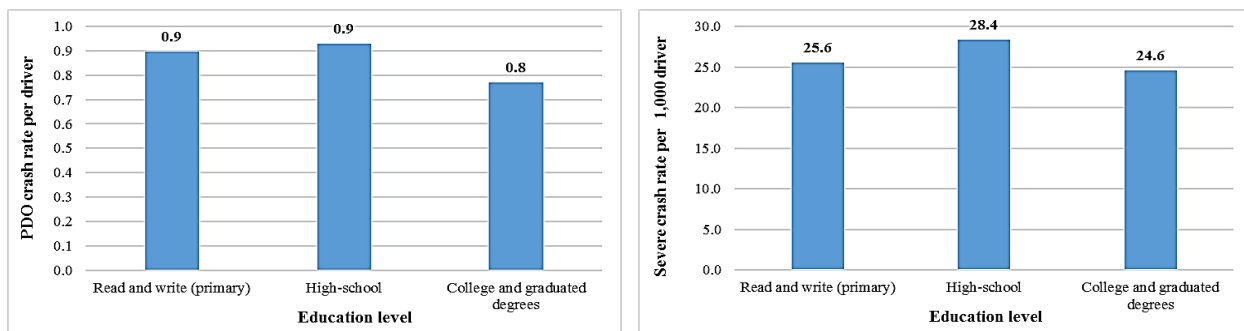Fig. 5: Crash rate in terms of driver's nationality.

Fig. 6: Crash rate in terms of driver's education level.

Figure 7 shows the crash rate in terms of the number of vehicles owned by the driver. It shows that the crash rate of the PDO crashes increases with increasing the number of the owned vehicles up to three vehicles after that the rate became stable. This results can be interpreted due to the high income of the driver. However, in case of the sever crashes, the crash rate seems not influences by the number of owned vehicles. Regarding the occupation (drivers against other occupation types), the workers as a driver (such as taxi drivers of driver in a company) seem have higher crash rate as shown in Figure 8. This is a logic result since drivers have more exposure of crashes due to high daily trips. Figure 9 shows the crash rate

with respect to the martial status of the driver. The results indicate that the martial status has no impact of the crash rate of the driver despite the crash rates of single drivers is a little higher which can be interpreted due to the high percentage of young drivers in AD (40% of licensed drivers aged less than 30 years old).
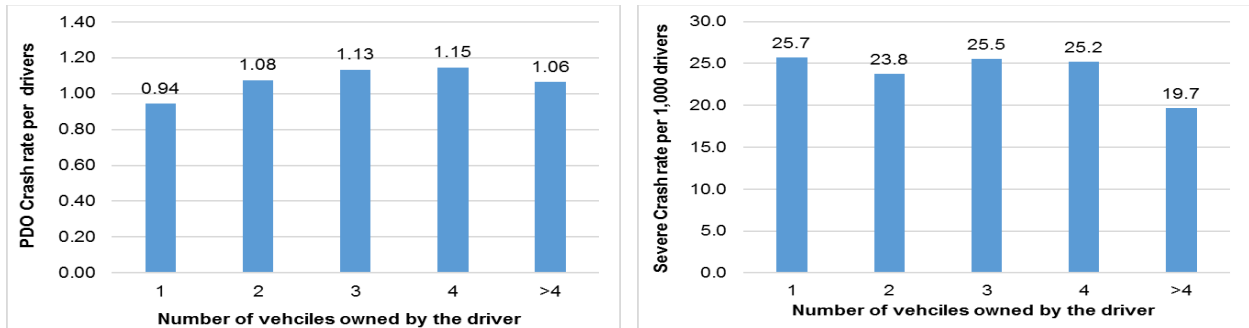

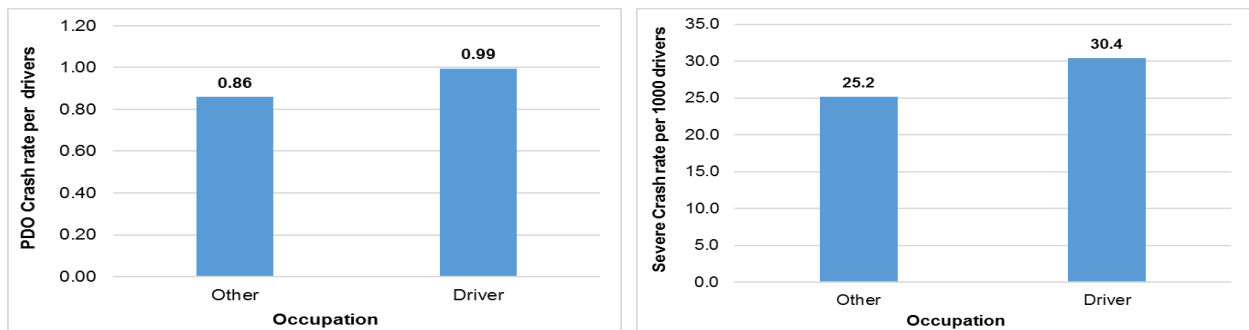Fig. 7: Crash rate in terms of number of vehicles owned by the drivers.

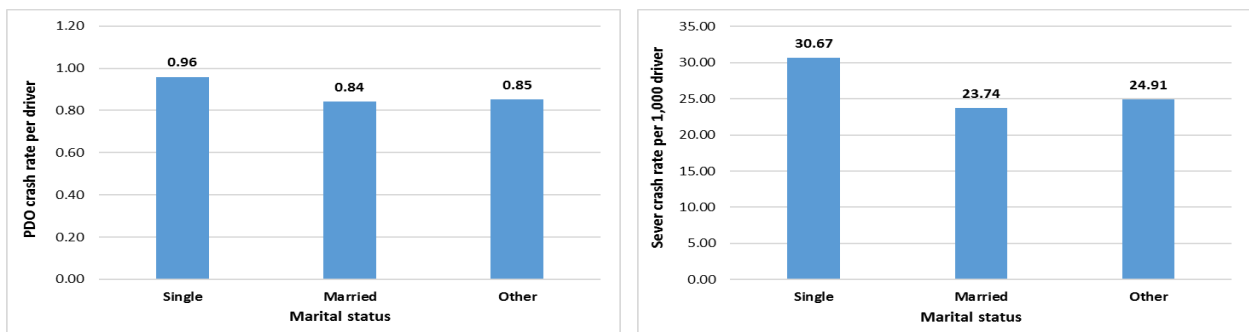Fig. 8: Crash rate in terms of the occupation type of the drivers.

Fig. 9: Crash rate in terms of the marital status.

4. Regression Model development

For further investigation of the interrelationship between the drivers' demographics, traffic violations and crash involvements, a regression model is developed to cover all of these variables in one model. Based on Highway Safety Manual (HSM-2010) the negative binomial regression modelling is considered as the best modelling approach to estimate the predictors variables of crash involvements. There for this model were applied in this study. The negative binomial regression model form to predict the total crash frequency is shown in Equation 2:

$$Ln\ Y = \beta 0 + \beta_1 X1 + \beta_2 X2 + \ldots + \beta_n X_n \tag{2}$$

Where, Y is the dependent variable (i.e. number of at-fault severe crashes in our case); X_1, X_2, ... X_n are the predictor variables; and β_0, β_1, β_2, ..., β_n are the regression coefficients. The SPSS statistical software package was employed to estimate the model using the customized negative binomial.

More variables regarding some types of violations that the drivers committed during the study periods were inserted in the examined variables package. there variables include speeding, tailgating, mobile usage and seat belt usage. Two models have been developed one for PDO crashes and the other of severe crashes estimations. the estimated parameters of the two models are shown in Table 1. The results show that the total number of the recorded violations of the driver is a significant variable in the occurrence of both PDO and severe crashes at a significant level of 95%. However, the education level of drivers is shown as a not significant variable in the model. On the other hand, the value and sign of the estimate and the odd ratio reflect the impact of the examined variable on the frequency occurrence of crashes. For example, the male driver increases the potential of involved in both PDO and severe crashes more than female (from the positive value of estimate). In addition, the odd ratio shows that male drivers have probably to be involved in PDO crashes by 1.137 times and involved in sever crashes by 1.369 times more than females. the same concept can be used to interpret all results shown in the table. other findings can be extracted from the results is that the Asians' drivers not significant in severe crashes but they are significant in PDO crashes involvements. Furthermore, the drivers who have mobile usage and tailgating violations are more likely to be involved in PDO crashes and such behaviour not significant in the occurrence of the severe crashes.

Table 1: Estimate parameters of the developed two models.

Variable			Parameter Estimates of the two models							
			PDO crashes model				Severe crashes model			
			Estimate	Chi-Square	Sig.	Odd ratio	Estimate	Chi-Square	Sig.	Odd ratio
(Intercept)			-.460	36.986	0.000	.631	-4.490	84.120	0.000	.011
Drivers' demographics		Total number of violations	.029	6053.471	0.000	1.030	.024	283.916	0.000	1.025
		Male drivers	.129	139.634	0.000	1.137	.314	29.618	0.000	1.369
		Age	-.005	103.150	0.000	.995	-.012	23.668	0.000	.988
		Experience	-.002	6.555	.010	.998	-.007	4.926	.026	.993
	Nationality	Emirati	.319	215.890	0.000	1.376	1.049	50.024	0.000	2.855
		GCC	.180	19.590	0.000	1.198	.860	15.680	0.000	2.363
		Arabian	.377	321.719	0.000	1.458	.603	16.861	0.000	1.827
		Asian	.232	113.186	0.000	1.261	.229	2.266	.132	1.257
	Education	Primary	.126	3.308	.069	1.135	.545	1.453	.228	1.725
		High School	.122	3.099	.078	1.130	.610	1.820	.177	1.841
		Graduated	.069	.969	.325	1.071	.506	1.244	.265	1.659
Not driver occupation			-.163	129.932	0.000	.850	-.632	79.649	0.000	.532
Number of owned vehicles			-.013	18.947	0.000	.987	-.082	28.440	0.000	.921
Violation types		Speeding	.022	2201.801	0.000	1.079	.017	86.163	0.000	1.083
		Mobile usage	.079	139.247	0.000	1.082	.002	.004	.952	1.002
		Seat belt	.035	122.492	0.000	1.035	.013	.846	.358	1.013
		Tailgating	.124	81.677	0.000	1.132	.317	37.808	.000	1.373

5. Conclusion

This paper investigated the interaction between the demographics characteristics of the drivers and their involvements of traffic rule violations and crashes. Both the property damage only PDO and severe (i.e., any crash resulted in at least one injury or fatality) crashed are separately investigated. Historical records of about 608,611 individual drivers during eight years were employed in the analysis. Crash rate of groups of drivers based on their demographics and violations were calculated. the analysis showed that crash rates of PDO and severe of drivers significantly increases with increasing the total number of the driver's violations. In addition, the demographic characteristics of the drives have a significant impact on the value of the crash rates. Male, young (18-24 years) older drivers (greater than 65 years), local nationality, high number of owned vehicle drivers, low educated and drivers who work as a driver have higher crash rates compared to others.

In order to consider the inter relationships among the tested variables, a model to estimate the variables that significant affect the frequency occurrence of the crash per driver has been developed by applying the Negative Binomial Regression approach. The results showed that some variables are significant in the occurrence of the both types of crashes such as number of violation records, drivers' age, gender, experience, occupation, number of owned vehicle by the driver, speeding violations, and tail gating violations. However, some variables are significant in the occurrence of the PDO crashes and not

significant in severe crashes such as Asian nationalities, mobile usage and seat belt violation types. in addition, the education level are not significant variables in the occurrence of the both types of crashes.

References

[1] J. R. Treat, N. S. Tumbas, S. T. McDonald, D. Shinar, R. D. Hume, R. E. Mayer, R. L. Stansifer, N. J. Castellan, "Tri-level study of the causes of traffic accidents," National Highway Traffic Safety Administration NHTSA, Report: DOT HS-805-099, 1979, [Online]. Available: http://ntl.bts.gov/lib/47000/47200/47286/Tri-level_study_of_the_causes_of_traffic_accidents_vol__II.pdf

[2] NHTSA, National Highway Traffic Safety Administration "Critical reasons for crashes investigated in the National motor vehicle crash causation survey," 2008, [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

[3] K. Rumar, "The basic driver error; Late detection," *Ergonomics*, vol. 33, pp. 1281-1290, 1990.

[4] L. Aberg, P. Rimmo, "Dimensions of aberrant driver behaviour," *Ergonomics*, vol. 41, pp. 39-56, 1998.

[5] J. Reason, A. Manstead, S. Stradlings, J. Baxter, K. Campbell, "Errors and violtions on the roads: A real distinction?," *Ergonomics*, vol. 33, pp. 1315-1332. 1990

[6] R. C. Peck, R. S. McBride, R. S. Coppin, "The distribution and prediction of driver accident frequencies," *Accident Analysis and Prevention*, vol. 2, pp. 243-299, 1971.

[7] J. R. Stewart, D. J. Campbell, "The statistical association between past and future accidents and violations," T*he University of North Carolina, Chapel Hill*, 1972.

[8] A. M. Gerbers, "Strategy for Estimating Drives Accident Risk in Relation to California\s Negligent-Operation Point System," California Department of Motor Vehicles, Research and Development Section*,* Report No. CAL-DMV-RSS-99-183, 1999.

[9] M. A. Gerbers, R. C. Peck, "Using Traffic Conviction Correlates to Identify High Accident-Risk Drivers," California Department of Motor Vehicles, Research and Development Section*,* Report: CAL-DMV-RSS-00-187, 2000.

[10] Y. Alver, M. C. Demirel, M. M. Mutlu, "Interaction between Socio-demographic Characteristics: Traffic Rule Violations and Traffic Accident History for Young Drivers," *Accident Analysis and Prevention*, vol. 72, pp. 95-104, 2014.

[11] G. Daigneault, P. Joly, J. Y. Frigon, "Previous convictions or accidents and the risk of subsequent accidents of older drivers," *Accident Analysis and Prevention*, vol. 34, no. 2, pp. 257-261, 2002.

[12] M. A. Gebers, R. C. Peck, "The identification of high-risk older drivers through age-mediated point systems," *Journal of Safety Research,* vol. 23, pp. 81-93, 1992.

[13] Y. Nishida, "Analyzing accidents and developing elderly driver-targeted measures based on accident and violation records," *International Association of Traffic and safety Science, IATSS Research,* vol. 39, pp. 26-35, 2015.

[14] S. Chandraratna, N. Stamatiadis, A. Stromberg, "Crash involvement of drivers with multiple crashes," *Accident Analysis and Prevention,* vol. 38, pp. 532-541, 2006.

[15] J. C. F. Winter, D. Dodou, "The Driver Behavior Questionnaire as Predictors: A Meta-Analysis," *Journal of Safety Research*, vol. 41, pp. 463-470. 2010.

[16] J. C. F. Winter, M. Kyriakidis, D. Dodou, R. Happee, "Using Crowd Flower to Study the Relationship between Self-Reported Violations and Traffic Accidents," *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, 2015.

[17] S. Obeid, V. Gitelman, O. Baron-Epel, "The Relationship between Social Capital and Traffic Violations: Israeli Arabs as a Case Study," *Accident Analysis and prevention,* vol. 71, pp. 273-285, 2014.

[18] J. L. Machado-León, J. de Oña, R. de Oña a, L. Eboli, G. Mazzulla, "Socio-economic and driving experience factors affecting drivers' perceptions of traffic crash risk," *Transportation Research Part F,* vol. 37, pp. 41-51, 2016.

[19] M. McDonald, R. Ingham, R. D. Hall, G. W. P. Rolls, "Accident risk and behavioural patterns of younger drivers," *Hampshire, UK: AA Foundation for Road Safety Research*, 1991.

[20] A. Mathews, "Biases in emotional processing," In: J. M. Jenkins, K. Oatley, N. L. Stein (Eds.), Human Emotions: A reader, Oxford: Blackwell publisher, 1998.

[21] J. A. Farrow, P. Brissing, "Risk for DWI: A new look at gender differences in drinking and driving influences, experiences and attitudes among new adolescent drivers," Health *Education Quarterly*, vol. 17, pp. 213-221, 1990.