

Factors Affecting Crash Frequencies: A Negative Binomial Regression Based Analysis of Indus Highway, Pakistan

Rafi Ullah Khan¹, Jingbo Yin^{1,a} and Faluk Shair Mustafa¹

1 School of Naval, Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract. The increase in vehicular traffic have also increased the highway crash frequency with the passage of time. Improvements in highway safety is of vital importance as it could save vast life and monetary losses. The highway crash frequency analysis of major Pakistani highways is a subject less discovered and many important strategic and trade routes are not studied in this regard. This study is aimed to analyze the crash frequency and the prominent factors that cause these crashes on a 302 km section of Indus highway; one of the most important trade routes of the country. Eight years' data from 2011 till 2018 was arranged into 19 variables where the crash frequency is set as dependent variable, while the eighteen prominent causation factors as independent variables. The tool used for analysis was negative binomial regression being run in the SPSS software. The results indicate that the driver's behavior, understanding & risk recognition, negligence and law adherence have a significant effect on the crash frequency. Furthermore, highway crash frequency significantly increases with increase in highway segment lengths, number of lanes and lane widths. Similarly, the highway crash frequency significantly enhances when the light, pavement surface and climate condition gets deteriorated. The results of this study are of vital importance to government, transportation companies and general public in order to recognize the most important accident causing factors and devise the transport policies, rules and behaviors accordingly.

1 Introduction

The severity level of crashes are subject to the cumulative influences of various observed and unobserved factors [1]. These factors could either be engineering or non-engineering. The engineering factors includes geometric and structural design of the roadway, traffic operation & management and the pavement surface condition [2]. The non-engineering factors stands for the behavioral patterns observed among drivers, various other human factors, effects of environment, seasonal and diurnal variations.

An overview of the past literature shows that in analysis of the highway crashes, the most abundantly used variables are related to the engineering and design. While the effect of factors other than engineering, specifically that of driver behavior have been significantly ignored [3]. Even though the behavior oriented studies reflect that driver behavior is the most prominent contributory factor, it is still not properly incorporated into crash risk modelling [4]. The reason for this is the unavailability of a profound and reliable data collection method and source that could collect the data related to various trends in the

driver's behavior [5]. The occurrence of a crash could be attributed to the driver being under the effect of alcohol or drugs, fatigue and carelessness.

The lack of data in this regard could be attributed to the reason that it is not the responsibility of federal or state highway agencies to take into record the behavioral patterns of the drivers involved in each accident. The data on the speed, lanes, distractions, disparities, and risky behaviors of drivers could be well found in the police reports of each accident [6].

The other prominent factors that effects the intensity and frequency of highway crashes are the sudden changes in the geometries of these highways, hindrances on the road way and the prevailing climatic conditions at the site of crash [7]. The climatic or weather alterations includes the onset of rains, tempests, smog, fog, slick pavements and inadequate light situations. All these factors affect the driver's ability to avoid any crash and collision on the road [8]. Moreover, the failures associated with brakes, axels, tires and other perfunctory culpabilities can result in crashes.

^a Corresponding author: jingboyin@sjtu.edu.cn

The tremendous increase in the motorized vehicles, the presence of a mix of motorized and non-motorized traffic and the susceptible transport modes like motorcycles & rickshaws have increased the highway crashes manifold in Pakistan [9]. The country has to face around a hundred billion rupees of annual losses in terms of fatalities, injuries and communal deficits resulting from highway crashes [10]. The highway crashes result in more than thirty thousand fatalities and around half a million injuries every year. Studies suggest among the contributory crash factors, 60% are attributed to the human factor, 30% to environmental factors, while 10% are attributed to the various mechanical faults associated with vehicles [11]. A better understanding of the contributory crash factors is pivotal in formulating better safety processes and policies.

This study is one of the few attempts to relate the highway crash frequency to various contributory factors like roadway geometries, environmental aspects, time variation and various human behavioral patterns. This study will evaluate the connectivity between road crashes initiated by geometric and traffic factors like number, type and width of lanes, shoulders, medians, access points, U-turns, curves, speed limits and traffic percentages. While the other factors taken into account were the age, gender, behavior patterns of the drivers, specific time, weather, and season at the time of crash along the vehicle conditions.

2 Literature review

Better understanding of the various contributory factors to crashes can result in better policies and measures for their reduction and management. The geometry of a road segment has a significant effect on the frequency and severity of crashes occurring in that segment [12]. The per unit length assessment of traffic crashes using Tobit regression model indicates that the pavement quality plays a key role in the crash occurrence [13].

A study of the Florida highway states that among the human factors analysis, 94% of the fatal crashes are of the kind where drivers were under the effects of alcohol [14]. The analysis of highway accidents in Pakistan reveals that unskilled drivers, heavily overloaded vehicles, pavement surface conditions and the use of mobile phones while driving are among the significant contributory factors [15]. A study conducted through multiple linear regression indicates that among the road crashes in Jordan, pavement surface and lighting conditions have the least effect on crash occurrence [16].

The behavior of a driver can be elaborated by his professional & social values along proclivity towards safety and risk [17]. The way a person looks at a risk could be affected by his society and surroundings. It could be better understood by observing that some of the drivers observe and follow other drivers in the surrounding and community instead of following the highway features and rules [18]. The expectation which a community or specific group of people in the

surrounding keeps also affects the way driver perceives risk [19].

The effect of a driver's behavior on a crash in the count related models could be replaced by substitute variables like that for over speeding [20]. Same technique could be adopted for driving under the influence of alcohol and drug [21]. Though alcohol and drugs are a type of societal crimes, however they are highly linked to the highway crashes and must be considered.

In order to evaluate the highway crash frequencies and the various contributing factors, various research methods and tools have been used over the period of time. In the past these techniques were simple and based on linear regressions mostly, however with the passage of time the normality of error, residuals and the other issues associated with linear relationship between various dependent and independent variables were sorted out [22]. Therefore, to accommodate the multifaceted relationship between various variables, Poisson regression models were started using which were based on the consideration of exponential relationship between the involved variables.

However, the dispersion of the data which required the equality of mean and variance for Poisson regression proved to be a limitation. This issue has been sorted out by the introduction of Negative binomial regression, which is considered independent of the over and under-dispersion limits. Though negative binomial regression model has some limitations associated with the multivariate variables, however it is considered very strong and efficient tool in evaluating the crash, fatality, and injuries related statistical data [2]. However, the use of this technique to evaluate the highway crashes data of Pakistan is very limited and various studies needs to be conducted evaluating various factors that affects the occurrence of crashes in land, air and maritime transportation.

3 Methodology & data collection

The purpose of this study was to determine those geometric, human, weather and vehicle factors which played effective role in the occurrence of crashes at the selected segment of the road. The relationship between the crash occurrence frequency and the engineering and non-engineering factors have been analyzed using the negative binomial regression. For this study, a segment of the Indus highway was selected which is one of the most prominent and strategic highway of the country. This highway serves as a hinterland connectivity, as it connects the ports of the country to its major cities

However, it is a very lengthy highway with a total length of 1264 km, connecting Karachi port city to Peshawar, therefore a section of it was considered for this study which is between Peshawar and Dera Ismail Khan. The length of this section is 302 km and passes through various rural and urban areas. The selected section was divided into five segments between the prominent cities on the route. The number of lanes in this segment is mostly two, however when it passes

through urban areas, the number of lanes has been increased to four. The type of median varies from none to grassy, curbs and paved at places. It has a number of horizontal and vertical curves with various bridges and access points. Since the highway is passing through various rural and urban areas, the speed limits it has, are varying. Since it is a key source of goods transfer, the traffic through it has a huge portion of heavy vehicles.

The data about the crashes in a specific area are collected by the regional police as per their jurisdiction, which is in the form of a properly detailed lodged FIR (First Information Report). Therefore, this study is based on the data extracted from the individual FIRs of each crash accidents in a specific segment of road from the police station of concerned jurisdiction.

These FIRs provide details on the types of vehicles involved, number plates, details on the conditions of drivers in the accident having information about the age, addiction and route or duration of driving. It also provides information on the level of experience the involved drivers had, their license details and the information about the possible causes after investigation which could have initiated these accidents.

A total of 8 years' traffic data was taken from the year 2011 to 2018. Since crash number is a count variable, hence best analyzed through Poisson and Negative binomial regression models.

Since the length of road segment with two lanes is higher than road length with four lanes, therefore the number of crashes occurring at two lanes is also higher than that occurring at four lane section as represented in figure 1.

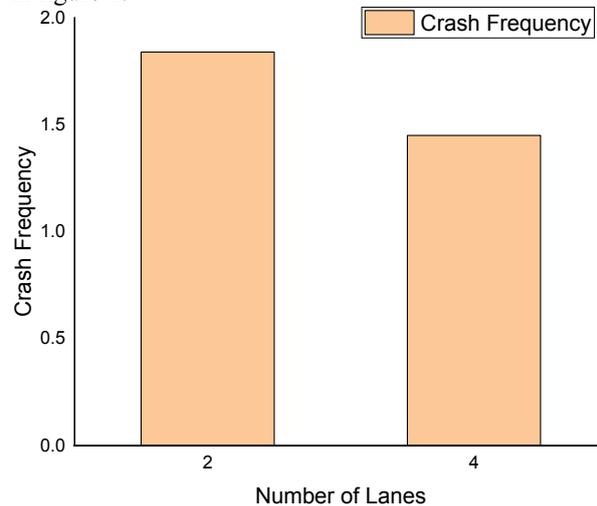


Figure 1. Graph between crash frequency and number of lanes.

The relationship between the crash frequency and speed limits at specific segments as per the statistical data under study shows that crashes at lower speed sections are more frequent than on high speed sections as depicted in figure 2.

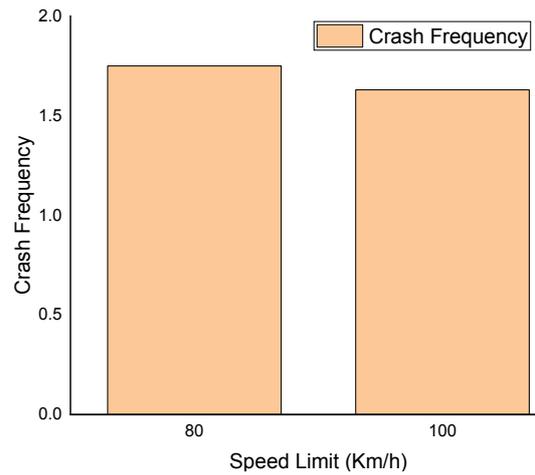


Figure 2. Graph between crash frequency and speed limits.

The pattern between the crash frequency and driver's behavior, climate effect, pavement condition and vehicle issues have been depicted in the figure 3.

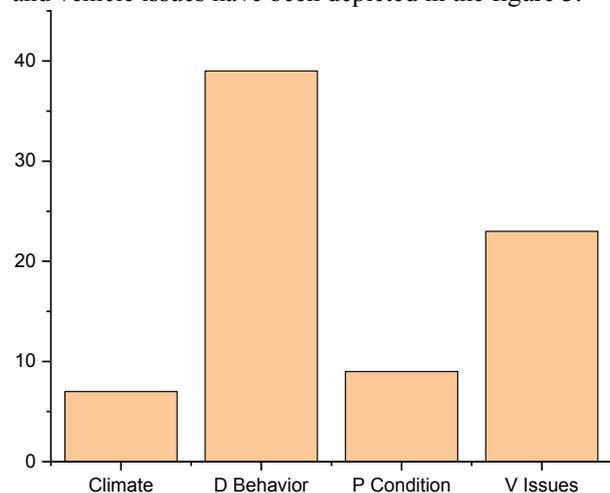


Figure 3. Graph between crash frequency and various factors.

However, the utility of Poisson regression model is limited by the ratio of mean to variance of the data which has to be equal to one. While this restriction is relaxed for the negative binomial regression. The data was found to be over-dispersed, the model suitable for this study is negative binomial regression (NBR). NBR actually is an addition of the Poisson regression, which have been provided with the measure for gamma distribution error [23]. Moreover, the selection of NBR was based on the performance of both dispersion and Vuong statistic parameters.

The mathematical expression to represent the negative binomial regression is given as;

$$Prob[Y = y_i | \epsilon] = \frac{\exp[-\lambda_i \exp(\epsilon)] (\lambda_i)^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots \quad (1)$$

$$\lambda_i = \exp(X_i \beta + \epsilon_i) \text{ for the } i\text{th observation; } (2)$$

Where, λ_i is the projected mean, β is a vector of venerable strictures, x_i is the factors selected as independent variables, and $\exp(\epsilon_i)$ is a gamma error term with mean 1.0 and variance α^2 . The adding of this constraint permits the variance to vary from the mean:

$$Var[y_i] = E[y_i] \{1 + \alpha E[y_i]\} = E[y_i] + \alpha E[y_i]^2 \quad (3)$$

The variable α , stands for the dispersion of the data, where if it gets equal to 0, the equation would become the Poisson regression model.

4 Results and discussions

In order to analyze the selected variables using NBR, first the mean, standard deviation, minimum and maximum values of the dependent and independent variables have been depicted in the table 1.

Table 1. Various statistical parameters of the selected variables

| No | Variables explanation | Mean | Std dev | Min | Max |
|----|--|-------|---------|------|-----|
| 1 | Dependent Variable: Crash frequency | 1.49 | 2.03 | 0.0 | 12 |
| 2 | Proportion of Unit trucks on a segment | 10.19 | 2.79 | 5 | 14 |
| 3 | Width of lanes in meters | 3.59 | 0.0782 | 3.65 | 3.9 |
| 4 | Segment wise lanes (Numbers) | 2.30 | 0.391 | 2 | 4 |
| 5 | Length of each segment | 1.11 | 0.789 | 0.15 | 6 |
| 6 | Access points in a segment | 2.09 | 2.87 | 0.0 | 14 |
| 7 | Urban and rural specification (1 and 0 respectively) | 0.44 | 0.55 | 0.0 | 1.0 |
| 8 | Allowed speed at each segment (km/hr.) | 87.89 | 14.87 | 80 | 100 |
| 9 | U-turns provided in a segment | 0.93 | 1.05 | 0.0 | 5.0 |
| 10 | Climate Impact | 0.82 | 0.99 | 0.0 | 1 |
| 11 | Priority Negligence | 2.24 | 1.03 | 3 | 9 |
| 12 | Indecorous reversing | 1.34 | 0.82 | 5 | 12 |
| 13 | Yielding negligence | 4.31 | 1.37 | 7 | 14 |
| 14 | Vehicle control failure while driving | 10.81 | 2.43 | 0.2 | 5 |
| 15 | Inappropriate turns | 8.95 | 1.63 | 0.0 | 6 |
| 16 | Wrong way or direction | 1.23 | 0.73 | 0.3 | 1 |
| 17 | Inappropriate Maneuvering | 3.78 | 0.058 | 3.0 | 6.0 |
| 18 | Light conditions | 2.96 | 0.89 | 3 | 4.5 |
| 19 | Pavement surface | 4.51 | 0.95 | 0.5 | 3 |

The results obtained after running the NBR analysis have been shown in table 2. Reference to this

table it could be seen that the number of single unit trucks are found to be significant with a t-value of 3.35 in causing highway crashes. This could be attributed to the reason that lighter trucks and vehicles have higher speeds as compared to the heavy trucks resulting in overtaking and increased crash risk. Similarly, the effect of a segment length on the crash frequency turns out to be significant with a t-stat value of 2.7. The results indicate that segments with longer lengths had contributed more to the occurrence of accidents as compared to shorter segment lengths. This could be attributed to the reason that a longer length of a road segment has potentially more exposure to the different traffic types and hindrances. These sections generally have higher number of U-turns, turns & horizontal curves and entrance roads. Due to the prevalence of similar road features, drivers have the inclination to do over speeding which also increase the crash risk.

Table 2. Results obtained from NBR analysis

| Variables explanation | Estimated Coefficient | t-Stat | p Value |
|--|-----------------------|--------|---------|
| Constant | -16.03 | -3.02 | 0.0004 |
| Proportion of Unit trucks on a segment | 0.19 | 3.35 | 0.0007 |
| Climate Impact | 0.54 | 1.27 | 0.0013 |
| Segment lane numbers | 0.89 | 2.90 | 0.0041 |
| Indecorous reversing | 0.62 | 1.59 | 0.5803 |
| Access points in a segment | 0.08 | 2.270 | 0.0244 |
| Vehicle control loss while driving | 0.28 | 1.55 | 0.0008 |
| Inappropriate Maneuvering | 0.60 | 2.44 | 0.0035 |
| U-turns provided in a segment | 0.08 | 1.04 | 0.3016 |
| Width of lanes in meters | 3.09 | 3.03 | 0.0021 |
| Priority Negligence | 0.78 | 2.54 | 0.0020 |
| Length of each segment | 0.29 | 2.70 | 0.0059 |
| Yielding negligence | 0.82 | 1.78 | 0.0033 |
| Urban and rural specification (1 and 0 respectively) | 0.69 | 1.58 | 0.2034 |
| Inappropriate turns | 0.72 | 1.63 | 0.0062 |
| Wrong way or opposite direction | 0.40 | 1.25 | 0.0006 |
| Allowed speed at each segment (km/hr.) | 0.02 | 1.21 | 0.2309 |
| Light conditions | 0.67 | 1.16 | 0.0051 |
| Pavement surface | 0.10 | 1.35 | 0.0064 |
| Dispersion parameter (a) | 0.51 | 4.709 | 0.0012 |

The climate effect on the occurrence of highway crashes is found significant with a t-value of 1.27, this is because in winters there is huge fog and smog

observed on this section of road. It results in very poor visibility and consequently traffic jamming. Similarly, the loss of control on vehicle and inappropriate maneuvering with t-values of 1.55 and 2.44 respectively were also found significant. The loss of control on a vehicle yields in a very high crash risk. The careless and inappropriate maneuvering puzzles the other drivers creating unsafe crossing & braking distances and subsequently resulting in crashes.

The width of lanes also turns out to have significant effect on the occurrence of highway accidents with a t-value of 3.03. The results reveal that wider lanes have higher tendency of resulting in an accident as compared to the narrow or constricted lanes. Though wider lanes mean more space, decision making time and better visibility, but still the results indicate otherwise. The potential reason is that drivers on lanes wider than normal tends to increase their speed which enhances the crash risk. Moreover, in a wider lane the drivers would see space along vehicles ahead in the same lane and will tend to cross or accompany these vehicles resulting in increased endangerment. The drivers on narrow lanes remains more careful with their speed and overtaking behaviors resulting in reduced crash risk.

The recognition, acceptance and understanding of the priorities in traffic while driving at a highway plays a very critical role in crash causation. The negligence in lane changing, passing, turning and yielding priority can result in misunderstanding and subsequently result in higher probability of crash occurrence. The t-values for priority and yield negligence were found to be 2.54 and 1.78 consequently with both of these variables being significant. Furthermore, making inappropriate turns or turns with wrong indicator also significantly increased the crash frequency with a t-value of 1.63. Similarly, driving in the opposite direction is very dangerous and results in very high probability of head-on and sidewise collisions having a 1.25 t-value. This puts the other drivers in jeopardy of making wrong maneuvers and turns to avoid the crash and hence result in collision with other vehicles.

Moreover, the number of lanes are also found to be significant in their effect on the crash frequency. It is attributed to the reasons that increased number of lanes would consequently result in increased traffic volumes, lane changing and higher speeds. All these reasons ultimately create circumstances in which higher crashes takes place. Similarly, urban areas are found significant in increasing crash frequencies with a t-value of 1.58. This is due to the increased volume of traffic, higher number of pedestrians, cycles, motorcycles and various carts. Another reasons which causes accidents are the encroachments in the urban areas which results in traffic jams and consequently impatience in drivers, resulting in close overtakes and higher crash frequency.

The light conditions prevailing at the road and the surface conditions of the road are also found significant. The poor light conditions results in poor visibility and wrong judgment of the size, speed and position of the other vehicles. While the poor surface

conditions can result in the loss of control over a vehicle and bursting, deterioration of tires and other parts consequently resulting in high vehicular crashes.

The speed limit, number of U-turns, number of access points and indecorous reversing though are very critical for traffic safety and crash frequency, but were found insignificant in the results of this study.

5 Conclusion

The number of traffic has increased manifold with the passage of time and so has the highway crashes. Highway crashes result in the loss of precious lives and property, hence enhancing the traffic safety is of remarkable prominence. The statistical data of the highway crashes and the factors due to which it occurs is best analyzed using Poisson or Negative Binomial Regression for reliable results.

Highway crashes have been a major concern in Pakistan for the authorities as it results in a significant annual life and monetary loss. Indus highway is one of the major trade routes of the country and its crash data have not been analyzed up till now as per the information of authors. The highway crash frequency data along its causation factors of this highway for a 302 km section from Peshawar to Dera Ismail Khan was taken and analyzed using NBR. The data was arranged as per the eighteen most important factors that affects the crash frequency and analyzed using SPSS software.

The results indicate that the crash frequency is highly subject to the length of specific road segments, number of lanes and the width of the lanes and an increase is observed in crash frequency with increase in these parameters. Similarly, the light, pavement and climate conditions also have a significant effect on highway crashes and the number of crashes increases as the condition of above parameters gets deteriorated. Similarly, the driver behavior, attitude and law abidance also plays a key role in the occurrence of highway crashes. Negligence, lower understanding & recognition, violation of rules and carelessness significantly increases the crash frequency.

References

1. M. R. R. Shaon, X. Qin, Z. Chen, and J. Zhang, "Exploration of contributing factors related to driver errors on highway segments," *Transportation Research Record*, vol. 2672, no. 38, pp. 22–34, 2018.
2. M. R. R. Shaon, X. Qin, A. P. Afghari, S. Washington, and M. M. Haque, "Incorporating behavioral variables into crash count prediction by severity: A multivariate multiple risk source approach," *Accident Analysis & Prevention*, vol. 129, pp. 277–288, 2019.
3. A. P. Afghari, M. M. Haque, and S. Washington, "Applying fractional split model to examine the effects of roadway geometric and traffic characteristics on speeding behavior," *Traffic injury prevention*, vol. 19, no. 8, pp. 860–866, 2018.
4. S. Washington and M. Haque, "On the commonly accepted assumptions regarding observed motor vehicle crash counts at transport system locations," 2013.

5. M. Shaon and R. Rahman, "Quantifying Rural Highway Safety Performance: Application of the Highway Safety Manual and Development of Mixed Distribution Statistical Models for Predicting Crash Frequency," 2015.
6. S. Box, "New Data from VTTI provides insight into cell phone use and driving distraction," *Virginia Tech Transportation Institute*, vol. 27, 2009.
7. M. R. Rahman Shaon and X. Qin, "Use of mixed distribution generalized linear models to quantify safety effects of rural roadway features," *Transportation Research Record*, vol. 2583, no. 1, pp. 134–141, 2016.
8. A. Montella and L. L. Imbriani, "Safety performance functions incorporating design consistency variables," *Accident Analysis & Prevention*, vol. 74, pp. 133–144, 2015.
9. I. Khan, A. Khan, F. Aziz, M. Islam, and S. Shafiqat, "Factors associated with helmet use among motorcycle users in Karachi, Pakistan," *Academic emergency medicine*, vol. 15, no. 4, pp. 384–387, 2008.
10. I. Ahmed, T. Islam, G. Ali, and M. M. Nawaz, "Pillion riders' cloth related injuries and helmet wearing patterns: a study of Lahore, Pakistan," *International journal of injury control and safety promotion*, vol. 23, no. 4, pp. 388–394, 2016.
11. D. Shinar, *Traffic safety and human behavior*. Emerald Publishing Limited, 2017.
12. A. Montella, L. Colantuoni, and R. Lamberti, "Crash prediction models for rural motorways," *Transportation Research Record*, vol. 2083, no. 1, pp. 180–189, 2008.
13. P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident rates using the random-parameters tobit model," *Accident Analysis & Prevention*, vol. 45, pp. 628–633, 2012.
14. L. K. Spainhour, D. Brill, J. O. Sobanjo, J. Wekezer, and P. V. Mtenga, "Evaluation of traffic crash fatality causes and effects: A study of fatal traffic crashes in Florida from 1998-2000 focusing on heavy truck crashes," 2005.
15. A. Ahmed, B. A. Khan, M. B. Khurshid, M. B. Khan, and A. Waheed, "Estimating national road crash fatalities using aggregate data," *International journal of injury control and safety promotion*, vol. 23, no. 3, pp. 249–254, 2016.
16. K. Jadaan, I. Al-Hyari, H. Naghawi, R. Ammourah, and Z. Al Nabulsi, "Traffic safety in Jordan: magnitude, cost and potential countermeasures," *J. of Traffic and Logistics Engineering*, vol. 1, no. 1, pp. 54–7, 2013.
17. P. M. Carter, C. R. Bingham, J. S. Zakrajsek, J. T. Shope, and T. B. Sayer, "Social norms and risk perception: Predictors of distracted driving behavior among novice adolescent drivers," *Journal of Adolescent Health*, vol. 54, no. 5, pp. S32–S41, 2014.
18. R. J. Schneider, A. Sanatizadeh, M. R. R. Shaon, Z. He, and X. Qin, "Exploratory analysis of driver yielding at low-speed, uncontrolled crosswalks in Milwaukee, Wisconsin," *Transportation research record*, vol. 2672, no. 35, pp. 21–32, 2018.
19. J. Moeckli and J. D. Lee, "The making of driving cultures," *Improving Traffic Safety Culture in the United States*, vol. 38, no. 2, pp. 185–192, 2007.
20. A. P. Afghari, S. Washington, M. M. Haque, and Z. Li, "A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk," *Analytic methods in accident research*, vol. 18, pp. 1–14, 2018.
21. M. Asbridge, J. A. Hayden, and J. L. Cartwright, "Acute cannabis consumption and motor vehicle collision risk: systematic review of observational studies and meta-analysis," *Bmj*, vol. 344, p. e536, 2012.
22. D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation research part A: policy and practice*, vol. 44, no. 5, pp. 291–305, 2010.
23. W. H. Greene and C. Zhang, "Econometric analysis (Vol. 5)," *Upper Saddle River, NJ: PrenticeHall*, 2003.