

# **Ridge Penalization-based Generalized Linear Model (GzLM) for Predicting Risky-Driving Index**

**Khadija B. Abdur Rouf**

Qatar Transportation and Traffic Safety Center (QTTSC)  
Qatar University  
2713, Al Hala Street, Doha, Qatar  
[kb1606107@qu.edu.qa](mailto:kb1606107@qu.edu.qa)

**Galal M. Abdella**

Department of Mechanical and Industrial Engineering  
Qatar University  
2713, Al Hala Street, Doha, Qatar  
[gmg5005@qu.edu.qa](mailto:gmg5005@qu.edu.qa)

**Khalifa N. Al-Khalifa**

Department of Mechanical and Industrial Engineering  
Qatar University  
2713, Al Hala Street, Doha, Qatar  
[alkhalifa@qu.edu.qa](mailto:alkhalifa@qu.edu.qa)

**Wael Alhajyaseen**

Qatar Transportation and Traffic Safety Center (QTTSC)  
Qatar University  
2713, Al Hala Street, Doha, Qatar  
[wjaseen@qu.edu.qa](mailto:wjaseen@qu.edu.qa)

## **Abstract**

Road traffic crashes remain one of the major causes of preventable death and injury worldwide. Human behavior is considered one of the main factors leading to such tragic losses. In this paper, we analyze the responses of an online survey questionnaire and identify the variables that are most likely to be correlated with individual driving behavior of drivers. Weights are allocated to nine risky-driving behaviors considered in the survey based on self-reported frequency of the driving behaviors the participants were involved in at the time of a recent traffic crash. Initially, weighted individual self-rated risky-driving behaviors are used to estimate the risky-driving index (RDI) for individual drivers. RDI is defined as a quantitative measure of a driver's risky-driving propensities based on basic profile and driving history. Finally, a standardized model for predicting a driver's RDI is proposed using Ridge penalization-based generalized linear regression with a standard error of estimate equal to 0.713. According to the model, female drivers have lower RDI compared to male drivers. Also, younger drivers have higher RDI than older drivers. Lastly, hours driven per day have more positive impact on RDI than the number of accidents or the driving experience of a driver.

## **Keywords**

Ridge penalization, Ridge regression, risky-driving index, RDI, driving behavior

## **1. Introduction**

Road traffic crashes remain a major concern for transportation and healthcare management agencies worldwide. According to World Health Organization (WHO)'s Global Status Report on Road Safety 2015, more than 1.2 million

lives are lost annually on world roads, and another 20~50 million people suffer from non-fatal injuries due to road traffic crashes (RTCs) (World Health Organization, 2015). Moreover, young males below 25 years are three times more likely to be killed in road crashes compared to young females. Consequently, 73% of all reported traffic crash deaths occur among young men. Besides, Treat (Treat, 1980) and Sabey and Staughton (Sabey & Staughton, 1975) found that in about 95% of the traffic crashes they studied in the US and the UK respectively, the road user was either the sole factor or the contributing factor. Likewise, Evans identified driving behavior to have a dominant role in road traffic crashes (Evans, 1996). Some key behavioral risk factors that influence crash involvement and severity are namely- speeding, drink-driving, and improper use or failure to use helmets, seat-belts, and child restraints (World Health Organization, 2015).

Speeding is a critical factor since the probability of crash increase as average traffic speed increase (Organisation for Economic Co-operation and Development, European Conference of Ministers of Transport, & OECD/ECMT Transport Research Centre, 2006). Consequently, the risk of death or serious injury in the event of a crash also increases at higher speeds (Rosén, Stigson, & Sander, 2011). Besides speeding, changing lanes, driving too close to the front vehicle, honking, passing on the shoulders, not wearing a seat-belt, etc. are some examples of aggressive or risky-driving behaviors (Jonah, Thiessen, & Au-Yeung, 2001; Shinar & Compton, 2004). On the other hand, Elvik R et al. (Elvik, 2003) reported that wearing seat belt reduces the risk of death or serious injury by about 45%~50% among driver and front passenger. As a result, implementation of mandatory seat-belt act has proven to increase adult seat-belt use and reduce traffic fatalities (Houston & Jr, 2006). Additionally, distracted driving is listed in the updated WHO road traffic injury fact sheet (2017) (WHO, 2017) as a contributing human factor to RTC. Slower braking reaction time, slower reaction to traffic signal, inability to remain in the correct lane, etc. are just some of the driving performance impairments caused by mobile use distractions. Such impairment puts the driver using a mobile phone at four times more risk of causing road crashes than the driver who is not using it (World Health Organization, 2015). Ghadban et al. (2018) investigated the comprehensiveness of traffic signs by drivers and the impact of driver characteristics gender, nationality, age, language, and educational level.

Thus, this paper aims to develop a regression-based model to predict risky-driving index (RDI) of individual drivers. The RDI is a new term used in this paper which, by definition, is a quantitative measure of a driver's risky-driving propensities based on the driver's basic profile (age and gender), driving characteristics (driving experience in years and hours driven per day), and traffic crash history (number of accidents in the past 3 years). In a RDI scale ranging from 0 to 5, the higher a driver scores, the riskier the driver's driving behavior is considered. Although this would essentially be a preliminary indication of a driver's high risky-driving tendencies, it could be used to identify high-risk drivers and implement necessary precautions or counter-measures.

Hence, in this study, a preliminary RDI is initially calculated using individual self-rated survey data collected from participants affiliated with Qatar University (see Section 3.5). The survey data includes the drivers' profile, driving behaviors, and traffic crash history. Finally, a prediction model is proposed to estimate the RDI of any driver (see Section 3.8). The rest of this paper is organized as follows: in Section 2, the methodology used for this study is explained. Results and discussions are covered in Section 3. Finally, the conclusions are discussed in Section 4.

## **2. Methodology**

### **2.1 Data Collection**

In many countries, police or national traffic crash data on driver's risky-driving behavior are either incomplete or difficult to obtain (World Health Organization, 2015). Thus, the data collection methodology used in this study was an online survey questionnaire in which relevant risky-driving behavior data was obtained without expending too many resources. Data used for this study was part of the data collected through the online "Driving Behaviors" survey conducted at Qatar University during the first month of the Spring- 2017 term.

### **2.2 The Design of Survey Questionnaire**

The survey questionnaire was designed to collect data such as the participants' gender (Male = 1; Female = 2), age in years (18~22 = 1; 23 ~30 = 2; 31~45 = 3; 46 ~ 65 = 4), driving experience in Qatar in years ( $\leq 2 = 1$ ; 3~5 = 2; 6~10 = 3;  $>10 = 4$ ), hours driven per day ( $< 1 = 1$ ; 1~2 = 2; 3~4 = 3;  $\geq 5 = 4$ ), number of traffic crashes they were part of in the last three years, driving behavior they were involved in at the time of the traffic crash selected from a list of nine risky-driving behaviors, and their self-rated driving behaviors for the same nine behaviors. For this study, emphasis was given on whether the participant was involved in a traffic crash and not on whether he/she was guilty. Similarly, the type or severity of the accident the participant was involved in was also not considered to be relevant, since the

risky-driving index is not proposed to be taken as an indication or degree of guilt. Rather, the RDI is a simple model that would not require hard to collect socially sensitive and legal information.

For the self-rated driving behavior section, the questionnaire was designed to yield unbiased behavioral responses by asking participants to rate the frequency of their “driving behaviors” instead of introducing the question to them as rating their “risky-driving behaviors.” It was designed such that the participants with higher rates would be considered as showing riskier driving behaviors. The participants were asked to choose from options such as always, often, sometimes, rarely, never, and not applicable. A scale of 1 to 5 (5 = Always; 4 = Often; 3 = Sometimes; 2 = Rarely; 1 = Never; Zero = Not applicable) was used to rate how frequently they said they practiced each of the following driving behaviors:

- Not wearing the car seat belt (DB1)
- Exceeding the speed limit (DB2)
- Driving in the emergency lane (DB3)
- Overtaking from the right side (DB4)
- Changing lanes without using the signs (DB5)
- Not keeping safe distance from the front vehicle (DB6)
- Eating or drinking while driving (DB7)
- Making or accepting phone calls (DB8)
- Reading or sending emails and text messages (DB9)

### 2.3 Statistical Model

The generalized linear model (GzLMs) is a commonly used form of linear regression statistical model when the response variable in a model does not satisfy the normality assumption (Cheng, Geedipally, & Lord, 2013; Sellers & Shmueli, 2009; Abdella *et al.* 2016). On the other hand, for estimating the coefficients of the model, many researchers prefer to use the maximum likelihood technique instead of the weighted least square technique. Likewise, it has also become the default technique to estimate coefficients in various statistical software packages. However, one issue with applying the maximum likelihood technique is the resulting instability in the model due to collinearity between the predicting variables. The penalization function such as the Ridge penalization is commonly added to the GzLMs for resolving such instability (Sellers & Shmueli, 2009; WHO, 2017).

Consequently, Ridge penalization-based generalized linear regression method is the statistical model used in this paper to estimate the risky-driving index (RDI) as a function of the driver’s profile. Equation 1 is the classical form of this model in which the response variable ( $y_i$ ) is continuous with more than one predictor variables ( $x$ 's).

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \quad i = 1, 2, \dots \quad (1)$$

where:  $x_{ij} \in \mathbb{R}^n$  is the  $i^{th}$  observation of the  $j^{th}$  model predictor,  $\beta_j$  represents the coefficients of the predictor variables, and the error term  $\varepsilon_i$  is assumed to be normally distributed with a mean  $\mu$  and a standard distribution  $\sigma$ . Now, based on the Ridge penalization-based generalized linear regression model proposed by Hoerl and Kennard (Hoerl & Kennard, 1970b), the modified values of the estimated  $\beta_j$  as a function of the Ridge-parameter ( $\lambda$ ) is found using (2). The Ridge-parameter  $\lambda$  is always a positive constant.

$$\hat{\beta}_{\text{Ridge}}(\lambda) = \operatorname{argmin}_{\beta} \|y - x\beta\|_2^2 + \lambda \|\beta\|_2^2 \quad (2)$$

### 2.4 Statistical Analysis

The STATGRAPHICS software was used to estimate the coefficients of the Ridge regression model and to choose a good value for the Ridge-parameter. The open source software R-studio was also used to cross-examine the obtained coefficients and the effect of the selected ridge-parameter. Both software generated similar results. However, STATGRAPHICS Centurion XVII was more versatile due to its built-in graphics and StatAdvisor tools. Besides, (1) and (2) were used in both cases.

## 3. Results and Discussions

### 3.1 Demographics of Participants

Table 1 shows the detailed sample proportion information ( $n = 500$ ) of the demographic variables selected for the modeling of the RDI. The proportion of male and female participants were almost equal. However, almost half of the

participants belonged to the youngest age range (18 – 22). In addition, only 38.0% of the participants had less than 3 years of driving experience in Qatar; whereas, about 62% of them had more than 3 years of driving experience. Again, almost half of those who participated said they drove for 1 – 2 hours daily.

Table 1. Proportion of selected socioeconomic variables of individual participants

Question	Answer	Proportion
Gender	Male	48.1%
	Female	51.9%
Age (years)	18 – 22	43.9%
	23 – 30	24.9%
	31 – 45	20.9%
	46 – 65	10.3%
Driving experience in Qatar (years)	≤ 2	38.0%
	3 – 5	33.4%
	6 – 10	14.1%
	>10	14.5%
Hours driven per day	< 1	16.3%
	1 – 2	45.1%
	3 – 4	32.6%
	> 5	6.0%

### 3.2 Analysis of the Demographics of Participants Involved in Traffic Crash

Only 33.8% of the 500 participants said they were involved in at least one recent traffic crash. Table 2 shows the selected demographics (gender and age) of those participants. Accordingly, males were more involved in traffic crashes than females for all age ranges. Whereas, males of youngest and oldest age range had a higher number of mean traffic crashes. On the other hand, the number of young drivers, less than 30 years old, involved in road crashes was greater than the number of older drivers. This indicates that young males are indeed at higher risk than any other group (World Health Organization, 2015).

Table 2. Demographics of participants involved in traffic crash in the last three years

Gender	Male			Female			
	Age (years)	≥ traffic crashes	Mean traffic crashes	Std. dev.	≥ traffic crashes	Mean traffic crashes	Std. dev.
	18 - 22	49.2%	1.79	1.22	37.5%	1.63	0.60
	23 - 30	64.4%	1.75	0.84	55.2%	2.03	1.07
	31 - 45	44.2%	1.65	0.70	32.3%	2.00	1.91
	46 - 65	30.8%	2.13	1.73	23.1%	2.00	1.00

### 3.3 Analysis of Driving Behavior during Traffic Crash

Participants involved in at least one traffic crash reported their driving behavior at the time of their most recent crash. The majority of them (36.5%) reported not keeping a safe distance between them and the vehicle in front. The second most reported behavior was not wearing a seat belt (17.5%) followed by over speeding, changing lanes without signs, or reading/sending emails/text messages (8% ~ 9.5%). 7% of them reported making calls on the mobile or overtaking from the right lane. Least reported behaviors were driving on emergency lane or eating/drinking (1.5% ~ 3.5%). See Figure 1 for more details. Besides, these results imply that all risky-driving behaviors lead to traffic crashes in varying degrees.

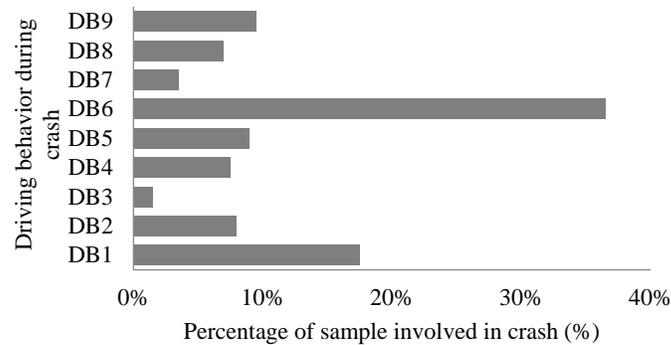


Figure 1. Frequency of the reported driving behavior of participants at the time of their most recent crash

### 3.4 Self-rated Driving Behaviors Analysis

Out of 500 participants, only 166 respondents had completely answered the self-rated driving behavior section of the questionnaire. The fall in the number of respondents could be attributed to the fact that it was the last section of the questionnaire. Also, compared to other sections, it required more thoughtful and time-consuming responses to questions related to socially and legally sensitive driving-behaviors. Hence, many participants just skipped the last section or didn't finish responding to all the nine questions. Based on the completed responses, value for each driving behavior (DB1 ~ DB9) of all 166 respondents were found ranging from 1 ~ 5 (Never ~ Always).

Consequently, an analysis of the distribution of the participants' self-rated driving behavior revealed that the majority of them had a value of DB1 = 1, that is, a low risk of not wearing seat-belt at most times. Similarly, the majority of them had low-risk value for DB3, DB4, and DB9. However, 0.5% ~ 14.7% of them reported that they would be often involved in all these driving behaviors. On the other hand, making calls on a mobile phone was the behavior that a small percentage of them (8.5%) reported as doing always (DB8 = 5). Besides, a small percentage of them (3.4% ~ 6%) reported always practicing behaviors such as exceeding speed limit, reading/sending emails/text messages, changing lanes without sign, not wearing seat-belt, or eating/drinking (see Figure 2 for more details of the distribution).

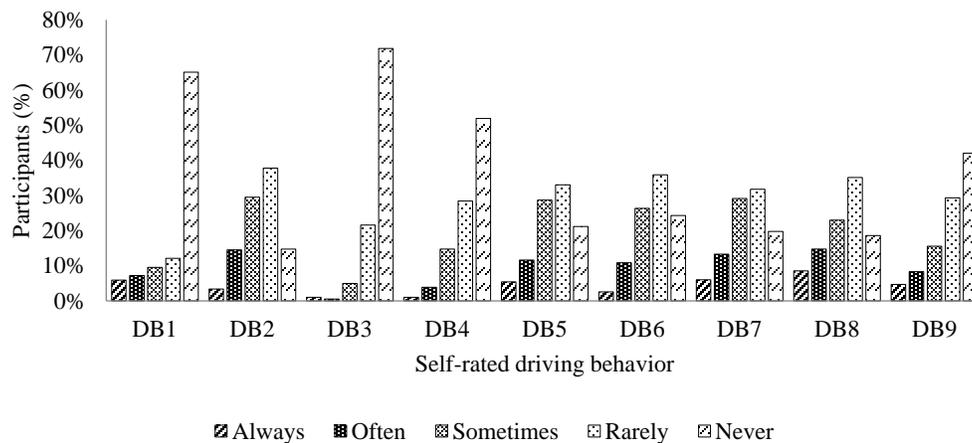


Figure 2. Distribution of participants' response to driving behavior practices

Besides, data of only 138 participants out of the 166 participants who completed the entire questionnaire was used for analysis and modeling of the RDI prediction model. Data of the participants who had less than or equal to 2 years of driving experience in Qatar was omitted from analysis. This was done because their responses to being involved in at least one traffic crash within the past three years could not be analyzed with the responses of those drivers who had more than 3 years of driving experience. Although the final sample size of 138 participants was much smaller compared to the initial sample size of 500 respondents, analysis of the data from this small sample was expected to provide more reliable results.

### 3.5 Individual Driver’s Risky-Driving Index (RDI)

Equation 3 is used to initially estimate the risky-driving index of individual participants based on their self-reported frequency of each driving behavior (see Figure 2) and selected weight (%) for each driving behavior. In addition, (4) shows the selected weights for each driving behavior. These selected weights, on the other hand, were assigned based on the varying percentage of the sample involved in traffic crash for each of the driving behaviors respectively (see Figure 1). In other words, summation of the selected values of these weights was equal to 100%.

$$RDI = (w_1DB_1 + w_2DB_2 + w_3DB_3 + w_4DB_4 + w_5DB_5 + w_6DB_6 + w_7DB_7 + w_8DB_8 + w_9DB_9) \times \frac{1}{100} \quad (3)$$

$$RDI = (17.5DB_1 + 8DB_2 + 1.5DB_3 + 7.5DB_4 + 9DB_5 + 36.5DB_6 + 3.5DB_7 + 7.0DB_8 + 9.5DB_9) \times \frac{1}{100} \quad (4)$$

### 3.6 Selecting Ridge-parameter

Ridge-parameter introduces a small amount of bias which controls the coefficients and reduces variance and prediction error of the GzLM when the independent variables are strongly correlated. The Ridge-parameter for the DBI prediction model was selected based on Ridge-traces (Hoerl & Kennard, 1970a). The variables considered in our analysis such as gender ( $x_1$ ), age ( $x_2$ ), driving experience in Qatar ( $x_3$ ), number of hours driven per day ( $x_4$ ), and number of accidents in the past 3 years ( $x_5$ ) were of quantitative nature, and they had varying rating scales and units of measurement. Hence, we chose the Ridge-parameter based on the Ridge-traces for the standardized regression coefficients instead of the general regression coefficients so that the prediction model will be valid for any units of measurement for the chosen variables. Figure 3 shows the impact of increasing the value of Ridge-parameter from 0 to 15 on the standardized coefficients of the GzLM. As predicted, the Ridge-parameter initially results in a dramatic change of the standardized coefficients, but as its value increases the estimates of the standardized coefficients begin to change slowly. This seems to happen when the value of Ridge-parameter gets close to 6.

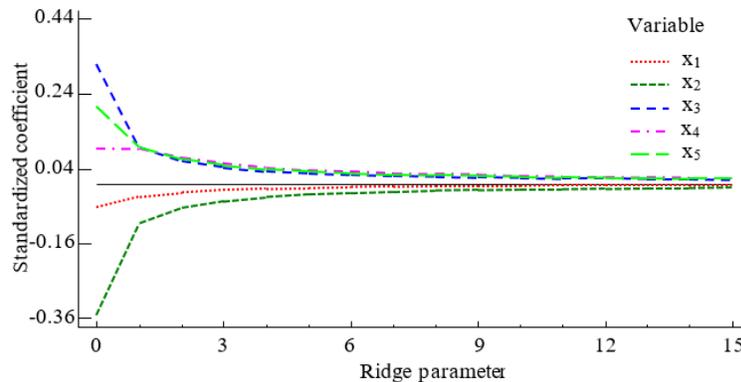


Figure 3. Ridge Trace for RDI

Figure 4 shows the measure of inflation of the estimated coefficients due to increasing value of Ridge-parameter from 0 to 15 when the independent variables are not correlated. Similar to standardized coefficients, the variance inflation factors also dramatically change at first but then change slowly and become stable around Ridge-parameter equal to 6. Also, we included all the initially selected independent variables in the GzLM, since the variance inflation factor for all them were less than 10.

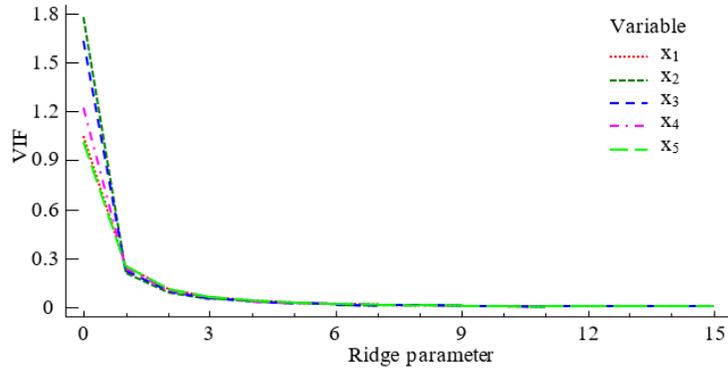


Figure 4. Variance Inflation Factors for RDI

Hence, based on visual inspection of the Ridge-traces generated by STATGRAPHICS Centurion XVII (see Figure 3 and Figure 4), Ridge-parameter was selected to be equal to 6, since this is where both the standardized coefficients and variance of inflation became stable. This ensures a GzLM that has low variance and prediction error.

### 3.7 Prediction Model for Driver’s Risky-Driving Index (RDI)

Equation 5 is the proposed general model for predicting driver’s RDI based on the classical model (1). The regression coefficients found by the ordinary least squares method and with Ridge-parameter  $\lambda = 0$  were the same as expected theoretically.

$$\text{RDI} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \quad (5)$$

where  $x_1$  = gender,  $x_2$  = age,  $x_3$  = driving experience in Qatar,  $x_4$  = hours driven per day,  $x_5$  = number of accidents in the past 3 years, and RDI = risky-driving index of a driver.

The abrupt and then steady change of the values of the coefficients and standardized coefficients of the regression model increasing with the value of Ridge-parameter  $\lambda$  are shown in Table 3 and Table 4 respectively. The nature of the rate of change of the coefficients is similar to the Ridge-trace in Figure 3.

Table 3. Comparison of coefficients of regression

$\lambda$	Coefficients of regression					
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
0	1.767	-0.092	-0.277	0.311	0.088	0.139
3	2.168	-0.026	-0.038	0.041	0.049	0.032
6	2.241	-0.015	-0.021	0.022	0.029	0.018
9	2.270	-0.010	-0.015	0.015	0.021	0.013

Table 4. Comparison of standardized coefficients of regression

$\lambda$	Standardized coefficients				
	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
0	-0.064	-0.351	0.322	0.094	0.205
3	-0.018	-0.048	0.042	0.053	0.047
6	-0.010	-0.027	0.023	0.031	0.027
9	-0.007	-0.018	0.016	0.022	0.019

### 3.8 Analysis of the Prediction Model

Equation 6 is the proposed standardized model for predicting driver's RDI with the selected Ridge-parameter  $\lambda = 6$ . By allowing a small amount of bias, the precision is greatly increased in this model. The standard error of estimate for this model is 0.713, and the mean absolute error is 0.543.

$$\text{RDI} = -0.010x_1 - 0.027x_2 + 0.023x_3 + 0.031x_4 + 0.027x_5 \quad (6)$$

The normal probability plot for the proposed RDI model (see Figure 5) shows that the residuals agree with the normality assumption. Also, the P-value for this model is greater than 0.05. This indicates that at the 95.0% confidence level there is no serial autocorrelation in the residuals shown in Figure 6.

### 3.9 Implications of the Prediction Model

The proposed standardized prediction model (6) has certain implications that agree with the conclusions of papers related to risky-driving behavior of drivers. The higher the value of a standardized coefficient, the more impact it has on the RDI. First, the negative value of the coefficients for gender and age indicates that young males are more susceptible to risky-driving behaviors than young female drivers. In other words, the lower the age, the higher the RDI value would become. Similarly, a male driver's RDI would be greater than a female driver's RDI. Other researchers have also reached similar conclusions regarding age and gender of drivers (Chen, 2009; Ghabban, Abdella, Al-Khalifa, Hamouda, & Abdur-Rouf, 2017; Machin & Sankey, 2008; Vassallo et al., 2007). Furthermore, the higher negative value of the standardized coefficient for age also indicates that it has more impact on RDI than gender. That is, regardless of the gender of the driver, a younger driver's RDI would be comparatively much higher than an older driver's RDI.

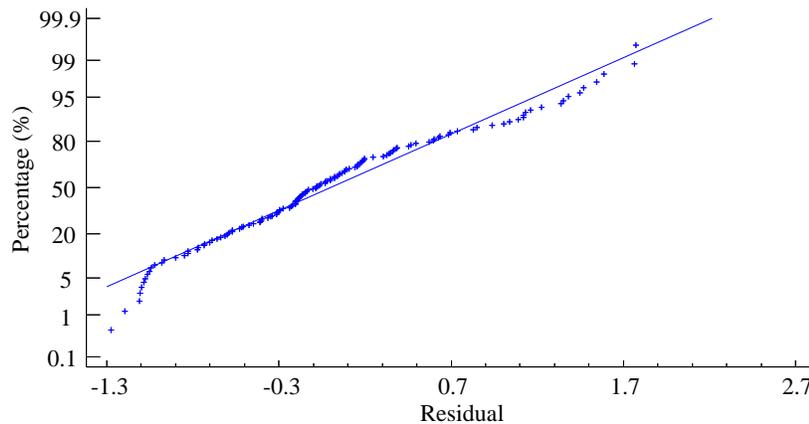


Figure 5. Normal probability plot for residuals

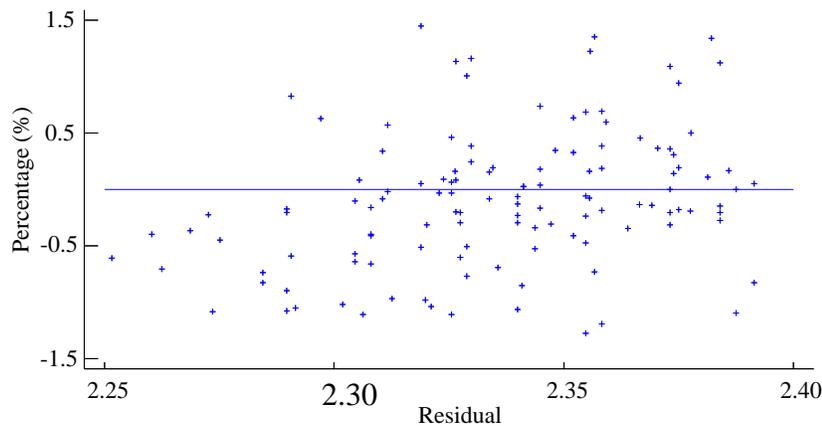


Figure 6. Residual scatter plot

Consequently, many researchers have studied the probable reasons behind this negative trend. For example, Vasallo et. al. studied risky-driving experiences and behaviors of young drivers aged 19-20 years (Vassallo et al., 2007) and found speeding by young males to be a major contributory factor to a high number of traffic crashes and injury rates. They also found that self-reported anti-social behavior of the young drivers also strongly contributed to their risky-driving behaviors. Senserrick & Haworth also came to similar conclusions that the more anti-social young drivers were, the greater were their speeding behaviors. Machin & Sankey, on the other hand, considered speeding to be the only factor contributing to risky-driving behavior (Machin & Sankey, 2008). They studied how personality and risk perception of young drivers influenced their driving behavior. They found young drivers aged 17-20 underestimated the risks involved with driving in various scenarios while overestimating their driving skills resulting in greater driving offenses. They also found that the driver's attitude about the society's acceptability of risky-driving behavior such as speeding plays the strongest role in the driver's choice to speed or take risks. Hence, most of these researchers also found the young age of these drivers' to be more of a contributing factor to high risk than their gender.

On the other hand, the positive value of the coefficients for number of hours driven per day, number of accidents, and driving experience in the proposed model indicates that the more a driver drives per day, and the more accidents the driver gets involved in, and the more driving experience the driver has, the higher the DBI for the driver becomes. Among all factors contributing to risky-driving propensities of a driver in our model, the coefficient for hours driven per day is the highest. This means that the more a driver drives per day, the higher the RDI would become. It is important to note that the model considers driving more than 6 hours per day the same as driving 10 hours per day. That is, anyone who drive's more than 5 hours per day will have the highest DBI.

In this model, the high risk associated with long hours of driving could be attributed to fatigue, drowsiness, distractions, or simply more exposure for the driver. According to a research, fatigue reduces a driver's capability on a short-term basis which in turn may lead to risk-driving behavior (Petridou & Moustaki, 2000). Moreover, results of a study in which fatigue was considered as a contributory factor in traffic crashes revealed that 4% of all crashes reported by drivers at fault were caused due to fatigue (Sagberg, 1999). However, since fatigue nor any breaks taken by a driver in between the mentioned driving hours per day was measured in this model, the underlying reasons need additional exploration.

Furthermore, the coefficient for the number of accidents the driver was involved in the last three years has the second most positive value. This implies that the higher the number of accidents the driver was involved in, the higher the driver's DBI would become. This is somewhat in agreement with a study in which the researchers found that the liability of an accident does not depend on a driver's tendencies for making errors, but it depends mostly on a driver's self-reported inclinations for not following traffic safety rules (Parker, Reason, Manstead, & Stradling, 1995). Consequently, the higher the likelihood of accident will be for a driver whose inclinations may have had already resulted in involvement with a previous traffic crash.

Lastly, the driving experience has the least positive impact on the value of the standardized RDI model. In a study by McKenna and Horswill, they identified self-ratings of a driver's driving skills and the associated driving thrill to

affect the driver's risky-driving behavior (McKenna & Horswill, 2006). Hence, the more experienced the driver, the higher the perception of skills and excitement triggering them to take more risks while driving. Besides, the driving experience is not an indicator of the age of the driver nor a measure of exposure for the driver. That is, a driver might have many years of driving experience, but he/she may have driven lesser than a driver who is young but has driven many hours per day. Consequently, the exposure of the younger driver would be much more than the older driver with just more years of driving experience. This could be one of the underlying reasons for the lower positive value of the coefficient of driving experience compared to the higher positive value of the coefficient of hours driven per day.

Although the model agrees with most of the popular findings in the literature, more research is needed to understand fully the impact of the combined effect of all the contributory factors. The underlying causes behind the implications of this model also needs to be studied further. Besides, the factors used in this model are basic and could be used for any region. Nonetheless, some other factors could be added to further improve and customize the model for other regions.

#### **4. Conclusion**

To sum up, in this paper a Ridge penalization-based generalized linear regression model for predicting a driver's risky-driving index (RDI) is proposed. This model provides a simple yet unique perspective to understand a driver's risky-driving behavior propensities based on the driver's basic profile and driving history. It requires using standardized independent variables such as a driver's gender, age, driving experience, the number of hours driven per day, and accident frequency to find an estimation of a driver's RDI with a standard error of estimate equal to 0.713. Hence, the RDI could potentially be used as a tool to check performance of drivers for hiring, educating, licensing, law enforcement, and other purposes. However, the proposed model is based on data exclusive to Qatar and at best valid for regions and situations similar to it. Thus, cross-cultural validation is required to verify reliability of the proposed model for other scenarios.

#### **Acknowledgements**

The authors acknowledge the support of the Qatar Transportation and Traffic Safety Center, Qatar. We would also like to thank the participants without whom the study would not have been possible.

#### **References**

- Abdella, G. M., Alhajyaseen, W., Al-Khalifa, N. K., & Hamouda, S. A. (2016). Usage of Non-Linear Regression for Modeling the Behavior of Motor Vehicle Crash Fatality (MVF) Rate. *The Proceedings of the International Conference on Industrial Engineering and Operations Management*, Rabat, Morocco, April 11-13, 2017, Pages 1827-1834. ISBN: 978-098554976-3
- Chen, C. F. (2009). Personality, safety attitudes and risky driving behaviors-Evidence from young Taiwanese motorcyclists. *Accident Analysis and Prevention*, 41(5), 963–968. <https://doi.org/10.1016/j.aap.2009.05.013>
- Cheng, L., Geedipally, S., & Lord, D. (2013). The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. *Safety Science*.
- Elvik, R. (2003). Effects on Road Safety of Converting Intersections to Roundabouts: Review of Evidence from Non-U.S. Studies. *Transportation Research Record*, 1847(3), 1–10. <https://doi.org/10.3141/1847-01>
- Evans, L. (1996). Comment: The dominant role of driver behavior in traffic safety. *American Journal of Public Health*, 86(6), 784–786. <https://doi.org/10.2105/AJPH.86.6.784>
- Ghadban, N. R., Abdella, G. M., Alhajyaseen, W., Al-Khalifa, N. K., Hamouda, A. M., & Abdur-Rouf, K. B. (2017). A Real Case-Based Study Exploring Influence of Human Age and Gender on Drivers' Behavior and Traffic Safety. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing*, vol 597 (pp. 807–816). California: Springer International Publishing. Retrieved from [http://link.springer.com/10.1007/978-3-319-60441-1\\_77](http://link.springer.com/10.1007/978-3-319-60441-1_77)
- Ghadban, N. R., Abdella, G. M., Alhajyaseen, W., Al-Khalifa, N. K., (2018) "Analyzing the Impact of Human Characteristics on the Comprehensibility of Road Traffic Signs", *The Proceedings of the International Conference on Industrial Engineering and Operations Management, Bandung, Indonesia, March 6-8, 2018* Pages 2210-2219.
- Hoerl, A. E., & Kennard, R. W. (1970a). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, 12(1), 69–82. <https://doi.org/10.1080/00401706.1970.10488635>
- Hoerl, A. E., & Kennard, R. W. (1970b). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Houston, D., & Jr, L. R. (2006). Safety belt use and the switch to primary enforcement, 1991–2003. *American Journal*

- of. Retrieved from <http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2005.074385>
- Jonah, B. A., Thiessen, R., & Au-Yeung, E. (2001). Sensation seeking, risky driving and behavioral adaptation. *Accident Analysis and Prevention*, 33(5), 679–684. [https://doi.org/10.1016/S0001-4575\(00\)00085-3](https://doi.org/10.1016/S0001-4575(00)00085-3)
- Machin, M. A., & Sankey, K. S. (2008). Relationships between young drivers' personality characteristics, risk perceptions, and driving behaviour. *Accident Analysis and Prevention*, 40(2), 541–547. <https://doi.org/10.1016/j.aap.2007.08.010>
- McKenna, F. P., & Horswill, M. S. (2006). Risk taking from the participant's perspective: The case of driving and accident risk. *Health Psychology*, 25(2), 163–170. <https://doi.org/10.1037/0278-6133.25.2.163>
- Organisation for Economic Co-operation and Development, European Conference of Ministers of Transport, & OECD/ECMT Transport Research Centre. (2006). *Speed Management*. OECD.
- Parker, D., Reason, J., Manstead, A., & Stradling, S. (1995). Driving errors, driving violations and accident involvement. *Ergonomics*, 38(5), 1036–1048. <https://doi.org/10.1080/00140139508925170>
- Petridou, E., & Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *European Journal of Epidemiology*, 16(9), 819–826. <https://doi.org/10.1023/A:1007649804201>
- Rosén, E., Stigson, H., & Sander, U. (2011). Literature review of pedestrian fatality risk as a function of car impact speed. *Accident Analysis and Prevention*, 43(1), 25–33. <https://doi.org/10.1016/j.aap.2010.04.003>
- Sabey, B. E., & Staughton, G. C. (1975). *Ceste i mostovi. CESTE I MOSTOVI*. Retrieved from <https://trid.trb.org/view.aspx?id=46132>
- Sagberg, F. (1999). Road accidents caused by drivers falling asleep. *Accident Analysis & Prevention*, 31(6), 639–649. [https://doi.org/10.1016/S0001-4575\(99\)00023-8](https://doi.org/10.1016/S0001-4575(99)00023-8)
- Sellers, K., & Shmueli, G. (2009). A regression model for count data with observation-level dispersion. In *24th International Workshop on Statistical*. Retrieved from [https://www.researchgate.net/profile/Galit\\_Shmueli/publication/228671467\\_A\\_Regression\\_Model\\_for\\_Count\\_Data\\_with\\_Observation-Level\\_Dispersion/links/0c960519b9173b238c000000.pdf](https://www.researchgate.net/profile/Galit_Shmueli/publication/228671467_A_Regression_Model_for_Count_Data_with_Observation-Level_Dispersion/links/0c960519b9173b238c000000.pdf)
- Shinar, D., & Compton, R. (2004). Aggressive driving: An observational study of driver, vehicle, and situational variables. *Accident Analysis and Prevention*, 36(3), 429–437. [https://doi.org/10.1016/S0001-4575\(03\)00037-X](https://doi.org/10.1016/S0001-4575(03)00037-X)
- Treat, J. R. (1980). A study of precrash factors involved in traffic accidents. *HSRI Research Review*. Retrieved from <http://psycnet.apa.org/psycinfo/1981-13775-001>
- Vassallo, S., Smart, D., Sanson, A., Harrison, W., Harris, A., Cockfield, S., & McIntyre, A. (2007). Risky driving among young Australian drivers: Trends, precursors and correlates. *Accident Analysis & Prevention*, 39(3), 444–458. <https://doi.org/10.1016/j.aap.2006.04.011>
- WHO. (2017). Road traffic injuries. Retrieved May 11, 2017, from <http://www.who.int/mediacentre/factsheets/fs358/en/>
- World Health Organization. (2015). Global Status Report on Road Safety 2015. *WHO Library Cataloguing-in-Publication Data Global*, 1–56. <https://doi.org/10.1136/injuryprev-2013-040775>

## Biographies

**Khadija B. Abdur Rouf** is currently a Master student in the Department of Civil and Architectural Engineering at Qatar University, Qatar. Ms Rouf holds Summa Cum Laude Bachelor of Science degree in Civil Engineering from Abu Dhabi University, UAE. Since her graduation in 2014, she has been a lecturer at University of Information Technology and Sciences, Bangladesh. She has also worked at Qatar Transportation and Traffic Safety Center, Qatar University as a Graduate Assistant during 2016-2017. Additionally, she worked on traffic impact study for Katara Cultural Village, Doha. Her research interests include traffic safety and operation, driving behavior, and public transportation. She's a recipient of Diversity Merit Scholarship from 2010 to 2014, Overall Award of Excellence from College of Engineering, Abu Dhabi University in 2013, and Internal Student Grant Award from Qatar University in 2018.

**Galal M. Abdella** is an Assistant Professor at the Mechanical and Industrial Engineering Department in the College of Engineering, Qatar University, Qatar. He was awarded his Ph. D. degree in Industrial and Systems Engineering from Wayne State University, Michigan, USA. Dr. Abdella is the founder of the professional IEOM chapter in Qatar. His research area has been centered on utilizing mathematics and advanced statistical data analysis for high-dimensional data processing, modeling and simulating rare events, quality data modeling and analysis, and project resource

management. Dr. Abdella has significantly contributed as a member of organizing committee or as a speaker in several workshops and forums on strategic topics related to transportation and road safety.

**Khalifa N. Al-Khalifa** is the Dean of College of Engineering and the founding director of Qatar Transportation and Traffic Safety Center (QTTSC) at Qatar University, Qatar. He is also a Professor in the Industrial Engineering and Mechanical Engineering Department, Qatar University. He was also the head of Quality Reliability and Maintainability Center, Qatar University. He obtained his Ph. D. in Manufacturing Engineering from the University of Birmingham, UK, and his Master and Bachelor degrees from Wayne State University, Michigan, USA, and University of Washington, Seattle, USA, respectively. His research mainly focuses on quality and reliability engineering. He is a member of IIE. He is also the Vice President of American Society for Quality, Qatar Branch. His paper in IIE in 2013 was the feature article of IIE magazine. Dr. Al-Khalifa is the chair of ASQ Doha-Qatar Local Member Community and a member of Qatari engineering society.

**Wael Alhajjaseen** is currently an Assistant Professor at the Qatar Transportation and Traffic Safety Center in the College of Engineering, Qatar University. He received the M.S. degree in transportation engineering from the University of Jordan in 2005 and the Doctor of Engineering degree in civil engineering from Nagoya University in 2010. He received several regional and international awards for his contribution in the field of Traffic Engineering such as Yasoshima Yoshinosuke Prize provided by the Eastern Asia Society for Transportation Studies in 2009. He is specialized in traffic engineering, traffic safety (human factors, crash data, and causality, driving simulation), road user behavior analysis (motorized & non-motorized), intelligent transportation systems.