International Journal of Civil Engineering and Technology (IJCIET)
Volume 8, Issue 11, November 2017, pp. 452–465, Article ID: IJCIET_08_11_048
Available online at http://http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=8&IType=11
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

AN OVERVIEW AND PREDICTION OF MALAYSIAN ROAD FATALITY: APPROACH USING GENERALIZED ESTIMATING EQUATIONS

N. Danlami
Department of Civil & Environmental Engineering,
Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar Malaysia

M. Napiah
Department of Civil & Environmental Engineering,
Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar Malaysia

A. F. M. Sadullah
Department of Civil Engineering. Universiti Sains Malaysia, Penang, Malaysia

N. Bala
Department of Civil & Environmental Engineering,
Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar Malaysia

ABSTRACT

The global effect of road traffic accident has led to numerous efforts by stakeholders to improve on road safety records at different levels. Countries set fatality reduction targets to be achieved over a certain period of time. An efficient scheme for fatality reduction comes with the need to have a fair idea of what the figures will be in the future. Fatality prediction models are used to predict the likely number of fatality over a specified time period. For the Malaysian case, several models are available including Negative Binomial Regression Models, Smeed's Law Models, and ARIMA model. In this article, Generalized Estimating Equation (GEE) is used to estimate road fatality using selected exposure variables. Population, Road Length, Vehicles involved in crashes and Mobile cell subscription per 100 people are found to be significant in predicting annual road fatality. GEE with negative binomial distribution was found to be suitable for use in short-term road fatality prediction modeling. A new exposure variable is proposed, tested and found satisfactory. The model developed was found to be reasonable when compared against similar existing models.

Keywords: Road Accidents, Injury, Fatality, Road, Population, Vehicle

Cite this Article: N. Danlami, M. Napiah, A. F. M. Sadullah and N. Bala, An Overview And Prediction of Malaysian Road Fatality: Approach Using Generalized Estimating Equations, International Journal of Civil Engineering and Technology, 8(11), 2017, pp. 452–465
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=11

1. INTRODUCTION

Road traffic accident remains one of the issues that attract a lot of attention throughout the world. While safety is outlined as one of the cardinal objectives of modern transport systems, road traffic crashes occur every day in different parts of the world. These crashes often claim lives and cause millions of injuries out of which many victims are maimed for life. According to the United Nation decade of action on road safety 2011-2020, about 1.24 million people are killed on the world roads each year. The report also indicated that more than 50 million are injured annually. Road traffic accident is the number one cause of death for young people in the world [1]. It is also in the records, road accident is the major cause of death of young people mostly in developing countries. Injuries from road crashes are said to be the seventh cause of death in the world and the figures are expected to rise to sixth by 2030 based on the current trend [2]. Besides being a serious public health problem, road traffic accident is also reported by the World Health Organization to have an effect on the economic development of a nation.

Despite the alarming statistics, not much relative attention is given to road traffic accidents worldwide most especially in developing countries compared to other similar incidences like plane crash, railway accidents, natural disasters or war.

Efforts are been made to improve the way roads are designed, constructed and managed. Environmental and other relevant standards are developed at national, regional and international levels aimed at attaining traffic safety objectives. Governments and agencies throughout the world have been channeling resources toward road injury and fatality reduction. The planning and execution of such programs require a good understanding of the influencing factors so that proper resource allocations can be made.

This article presents an exploratory study on the significance of some variables on road fatality rates in Malaysia over a period of 36 years. The primary indicators in the context of this article are the risk and exposure factors which are considered as macroscopic variables [3]. These variables include road length available to the drivers, growth in overall car ownership and population. While on the other hand, the secondary variables constitute gross national income per capita (GNI), mobile cellular usage per 100,000 population, and GDP growth. The impact of each of the variables on road fatality is investigated using statistical means, and only those that are found very relevant are discussed. Short term fatality prediction models are formulated from the interaction of the variables. The models are intended to be useful in the planning of emergency road safety intervention programs. In addition, the models from this study will serve as a review to some of the existing fatality prediction models available in the country.

2. BACKGROUND

Road traffic accident poses a range of challenges to countries globally, it is said to be the leading cause of premature death [4] it is the ninth leading cause of death among all age groups and number one for age 15-29 throughout the world [2]. As stated earlier, the record at hand does not only present road crashes as a serious health issue but also as a huge economic problem. The age group 15-29 years who constitutes the majority of the casualties are supposed to be the center of the workforce in the economy of any country. In some instances, members of this age group are relied upon to provide to their families.

The United Nation Decade of Action for Road Safety program aimed at intervening in the raising global rate of fatality and injuries has resulted in many countries taking measures towards achieving the road safety objectives. According to IRTAD (International Road Traffic and Accident Database), the total number of fatality from road accident recorded in its 32 participating countries has declined by about 42% between the year 2000 and 2013. Reductions of up to 50% were recorded in Lithuania, Denmark, Slovenia and France while Spain and Portugal recorded the biggest reduction of about 70%. The problem, however, is that the reduction recorded in non-European countries is much lower than the general average.

Road deaths in low and middle-income countries account for more than 90% of the annual total [5]. Consistent record showed that young drivers stand a high risk of accident among the population of drivers in a country. In this regard, developing countries are more likely to have higher accident rate due to the high percentage of young people and high fertility rate compared to developed countries where older age group constitute the vast majority. Malaysia, in addition to having a high population of young people, the country also have a considerably large number of motorcycles.

2.1. Road Safety in Malaysia

Malaysia is a middle-income country. It is reported to be among the world top 20 economies and second after Singapore on the ease of doing business in South-east Asia [6]. However, as the economic activities grow it is accompanied by urbanization and motorization. In such instances, it is highly likely that car ownership will increase due to the improved economic status of the people and the corresponding need for mobility. Between 2004 and 2014 there is an increase of 82% in the number of vehicles registered in the country compared to 18% increase in the nation's population within the same period. Road accident has raised by about 42% within the last decade with the number of deaths for the year 2014 recorded at 6674. Road traffic fatality per 100,000 population for Malaysia stands at 24 compared to the global average of 17.4. This figure is considered very high and it is a cause for great concern. Over the years the rate of road accident has been on the rise as shown in Figure 1. But the number of injuries recorded has been on the decline since 1998. The gain in the declining number of injuries is, however, unjustified by the rising number of annual recorded deaths. The rise in the number of death has reached to an alarming figure of about 18 counts per day in 2014.

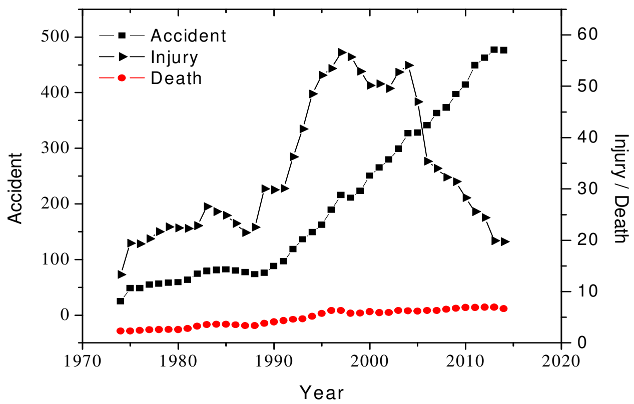

Figure 1: Road Safety Record 1974-2014

2.2. Road Fatality Modelling

Prediction of fatality resulting from road accidents is an important component of road safety management. Several models were developed in different countries, some of which are applicable globally and those with local outlook. Earlier in 1948 J.R Smeed suggested that the number of fatalities recorded in any country in any given year is related to the population of that country and the number of registered vehicles [7]. Many studies including [8-10] have been reported to have made an attempt to improve on Smeed's perspective by considering other variables. Some of the factors considered in these studies comprise of driver aggression, speed limit, vehicle kilometer traveled, income level, age, alcohol usage, and population. Forecasting fatality has become more interactive in term of variables and modeling skills, due to the availability of more tangible data and robust analysis tools compared to the 1940s.

2.3. Goal Setting and Fatality Reduction Target

Reduction in fatality rate is one of the key objective of road traffic fatality forecasting throughout the world. Therefore, depending on the country and circumstances, the prevailing factors are normally targeted. Marsden and Bonsall [11] recommended three approaches for goal-setting when forecasting for fatality reduction namely; model-based, under which the application of system interaction and transport user is proposed to determine variable response under different policy scenarios. Secondly, extrapolation and evidence-led judgment on the basis that not all indicators can be modeled. The last point is based on inspiration, under which, targets are set based on the best professional assessment of what is intended to be achieved. Model-based target setting was reported to be the most thorough [12].

2.4. Road Fatality Prediction Techniques

Several techniques have been used in the modeling of road crashes and fatality. For example, Chang and Mannering [13] investigated the significance of truck involvement in the number of fatalities using nested logit model. On another note, [14] used a logistic regression model to determine the independent influence of crash, driver and vehicle characteristics on fatality among drivers. MacLeod et.al [15] used logistic regression to model the factors associated with hit-and-run on pedestrians, in which the research concluded that some social factors including alcohol and invalid driver's license can be key factors. Similar research works that used the logistic regression model include [16-19]. However, it is important to note that logistic regression requires large sample size for a better result [20].

In situations where data is collected and framed over a time period, the prediction can best be made using time series technique. The most popular time series model used in fatality forecasting is the Auto Regressive Integrated Moving Average (ARIMA). ARIMA was introduced by Box and Jenkins [21] and have been used in long-term fatality forecasting. Adu-Poku et.al [22] modeled twenty years (1991-2011) data using ARIMA to investigate road fatality in Ghana. The study predicted a rising fatality rate over the next five years in the country. Similar findings were reported in China and Malaysia using the ARIMA model [23, 24]. However, [25] argued that, for a non-negative integer count data, for instance, road traffic accident, the use of time series models such as ARIMA may be unsuitable. This is due to the fact that the errors in the ARIMA model are assumed to be normally distributed and therefore, a model capable of dealing with a serial correlation of the variables should be used.

Accident data are discrete, random and only take a value of a non-negative integer. In some instances, the data for accident and fatality investigation is cross-sectional in nature. Over the years, the use of ordinary least square regression (OLS) in modeling accident and fatality have been discouraged [26-29].

Statistically, OLS assume a continuous dependent variable such that an observation can take a positive or negative decimal value. The reality, however, there is no fraction of accident or fatality and therefore other models like the Generalized linear models (GLM) are explored. GLM have been used mostly in fitting the relationship between accident and other variables. The most common forms of GLMs used are the Poisson regression and the Negative Binomial Regression (NBR).

2.4.1. Poisson and Negative Binomial Models

In Poison model, the dependent variable is assumed to follow a Poisson distribution. In this case, each of the observed accident or fatality n_i is drawn from a Poisson distribution that has a conditional mean (of observed fatality) λi with a vector Xi for expected number of fatality per year *i.* The probability density function (PDF) for an observation *ni* can be written as presented in Eq 1;

$$P(n_i|X_i) = \frac{EXP(-\lambda)\lambda^{ni}}{n_i!}, \; for \; (n = 0,1,2......) \tag{1}$$

Where; Poisson parameter ***(λi)*** is given as a function ***λi=EXP (βXi)*** in which Xi is vector explanatory variable and β is an estimable vector parameter. Poisson distribution has a log function as a canonical link function in which ***Xβ=ln(λ);***

$$\log(\lambda_i) = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + ...\beta_j x_{ij} \tag{2}$$

Where β's are the estimated regression coefficients and x's are corresponding independent variables. The regression coefficients represent the anticipated change in the log of the mean λi per a corresponding unit change in the independent variables x's.

Poisson regression assumes that the data has equal mean and variance **(E[ni] =VAR [ni])**. Such assumption is unrealistic with count data. This assumption has led to several attempts to improve on the model, such improvement includes the Zero-Inflated which is reported to give good model estimations [30]. Connected to that, Lord and Mannering [31, 32] gave a comprehensive comparison of some suitable models.

To address overdispersion, Poisson regression model was improved with a Gamma function [33]. Hence, the Poisson parameter can be re-written as; ***λi=EXP (βXi+εi)***

Where **EXP (*εi*)** with a mean λi equal 1 and a variance **α**, represents a corrective Gamma-distributed error term to make the mean differ from the variance. From this improvement, the Probability Density Function for negative binomial distribution with Γ as Gamma function can be presented as:

$$P(n_i) = \left(\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right)^{1/\alpha} \frac{\Gamma[(1/\alpha)+n_i]}{\Gamma(1/\alpha)n_i!} \left(\frac{\lambda_i}{(1/\alpha)+\lambda_i}\right)^{n_i} \tag{3}$$

Likelihood functions are used to justify unobserved factors in the observations with random parameters. The estimable parameter β in the Poisson distribution is then modified with a randomly distributed term φi as shown in Eq 4:

$$\beta_i = \beta + \varphi_i \tag{4}$$

2.4.2. Generalized Estimating Equation

Generalized Estimating Equations (GEE) is a semi-parametric method used in the estimation of GLM parameters with a likely unknown correlation between outcomes [34]. GEE provides a more general approach for the analysis of correlated responses between variables. When estimating the mean responses and analysis of within subject association structure, the GEE is found to be more flexible then other relevant techniques [35]. Besides addressing overdispersion, GEE also checks serial correlation in the data.

Crash frequency data repeatedly observed in one location over a given period of time is very likely to be correlated. It is therefore, necessary to account for temporal correlation when formulating a model using longitudinal data. If y_{in} donates the observed accident frequency recorded for a location *i* in a given year *n* , where *i =1* and *n* = 1, the coefficient β is estimated in GEE by working the quasi-score function shown in equation 5 as a differential function [36]

$$U_j(\beta) = \sum_{i=1}^{j} \frac{\partial \mu_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i) = 0 \qquad (5)$$

Where:

$U_j(\beta)$ = quasi-score function

Y_i and μ_i are vector of observed accident frequency at location *i*'s and vector of the anticipated value of accident frequency at same locations respectively

V_i = Y_i's covariance matrix specifying the type of temporal correlation of the observed accident frequency. V_i is given in equation 6 as follows;

$$V_i = \frac{A_i^{1/2} R_i(\alpha) A_i^{1/2}}{\phi} \qquad (6)$$

$R_i(\alpha)$ denotes the within panel correlation matrix replacing the identity matrix in GLM.

ϕ is a scale parameter, while;

$A_i^{1/2}$ is a diagonal matrix of N x N having variance function of $\boldsymbol{Y_i}$, $\boldsymbol{v(\mu_i)}$ as the k^{th} diagonal element

The correlation structure is determined by the value of the vector parameter α. For instance, since 36 year accident data is used in this work, the working correlation matrix will be a 36 x 36 matrix as shown in the Autoregressive matrix in Figure 2.

$$R_{36x36} = \begin{matrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & . & . & \alpha^{35} \\ \alpha & 1 & \alpha & \alpha^2 & \alpha^3 & . & . & \alpha^{34} \\ \alpha^2 & \alpha & 1 & \alpha & \alpha^2 & . & . & \alpha^{33} \\ \alpha^3 & \alpha^2 & \alpha & 1 & \alpha & \alpha^2 & . & \alpha^{32} \\ \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 & \alpha & . & . \\ . & . & . & . & . & 1 & \alpha & . \\ . & . & . & . & . & . & 1 & \alpha \\ \alpha^{35} & . & . & \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \end{matrix}$$

Figure 2 Autoregressive structure correlation matrix

There are basically four approaches to correlation structure in GEE [36]. These include Independent $R_i(\alpha)$, Exchangeable $R_i(\alpha)$, Unstructured $R_i(\alpha)$ and Autoregressive *(AR)* $R_i(\alpha)$. In the AR $R_i(\alpha)$, the assumption is that the correlation between any two observed accident frequencies in two consecutive years is stronger than that between observed frequencies in non-successive two year.

A preliminary run of the data on SPSS 23 was first carried out to determine the best correlation structure. Out of the four mentioned earlier, AR is found to be more adequate.

Two different distribution i.e. Negative binomial and Poisson Log-linear are used afterwards in the model formulation and the best models are selected and discussed.

In AR, α is a vector and Pearson residuals r is used in estimating the correlation [37]. The structure is shown in Eq(7). As shown in figure 2, since the value of the vector parameter α is smaller than unity, the correlation will decrease with increasing gap between the years in which the accident frequencies are observed.

$$\alpha = \frac{1}{\phi}\left[\sum_{i=1}^{n}\left(\frac{\sum_{t=1}^{n_i-0} r_{i,t} r_{i,t+0}}{n_i}, \ldots, \frac{\sum_{t=1}^{n_i-k} r_{it} r_{i,t+k}}{n_i}\right)\right] \tag{7}$$

3. DATA DESCRIPTION AND PRESENTATION

Data from 1974 to 2010 was used to establish the long-term significance of the selected variables on road fatality. A statistical summary of the data is given in Table 1. The data came from three main sources namely; (i) Transport Statistics Malaysia published annually by the Ministry of Transportation Malaysia. (ii) Road Accident Statistics Report: This is also published annually by the Royal Malaysian Police. The road Safety Annual Report prepared by Malaysian Institute of Road Safety Research (MIROS) is a summary and interpretation of the data collected from this source. And (iii) The World Bank [38]. Four of the variables used in the models (namely: Population, Road length, Vehicle ownership, and the number of Vehicles involved in crash) retained their standard definitions. While Mobile subscription and VI/RL are defined in the context of this study as follows;

- **Mobile Cellular Subscription (Mc)** represents the number of mobile cell subscription per 100 people in the country as obtained from the World Bank records[38].
- **VI/RL:** This is a transformed exposure variable introduced for the purpose of this study. It represents the number of vehicles involved in crashes per kilometer of public road in any given year.

Preliminarily, a scatter plot of all the variables was carried out to physically identify correlation between the variables. Further checking for linear dependence between the candidate variables was done using the Pearson correlation coefficient. Based on the correlation result, P, Mc, RL and VI/RL were selected to be used in the modeling.

Table 1 Statistical Summary of Data

Variable	N	Minimum	Maximum	Mean	Std. Deviation	Dispersion (μ-σ)
Road Accidents	36	24581	477204	198184	144668	53516
Road Injuries	36	13332	56574	33432	13232	20200
Road Death	36	2303	6917	4811	1632	3179
Population	36	10434592	30300000	20092226	6343180	13749046
Vehicles Registered	36	1090279	25101192	9045467	7110234	1935233
Vehicle Involved	36	39056	921232	355534	274058	81476
Road Length (km)	36	11161	135714	59563	32246	27316
Vehicle Ownership	36	1	10	3	2	1
Mobile cells Subscriptions/100people	36	0	149	33	49	-16
GDP Growth Annual	36	-7	12	6	4	2
GNICurrentUSDx10E09	36	9	327	96	93	4

4. RESULTS AND DISCUSSION

In GEE models with loglink functions, temporal correlation among observed accident frequencies and the nonlinear effect of the independent variables were taken into consideration. The correlation structure was computed with the assumption that measurement was equally spaced for all the variables. Two different models were developed with their subjects selected after preliminary check for correlation and linear dependence using scatter plot and Pearson's correlation respectively.

As earlier mentioned, four correlation structures were compared in the GEE model for both the P-VI/RL and VI/RL-Mc model combinations. The result suggested that the AR correlation structure is most suitable having lower QIC (Quasi-Akaike Information Criteria).

The first model having a variable combination of Population and VI/RL shows that both variables are statistically significant at 95% confidence level. The relationship indicates positive estimate for population, thus suggesting an increase in the annual fatality figures with increasing population of the country. This relationship was upheld for both Negative Binomial and Poisson Log-Linear distributions. Based on the analysis done, the annual number of death is expected to rise with a corresponding increase in population independent of other contributing model variables. For the Malaysian case, a similar finding was presented by [39, 40].

Initial scatter plot of the data revealed vehicle ownership (motorization) have similar effect as population on annual fatality rate. Vehicle ownership and population appear to be on the constant rise in Malaysia for the past 40 years. However, studies show that reduction in fatality rate could be achieved if the two variables are controlled [41].

Conversely, road fatality figures are shown to decrease with an increasing number of vehicles involved in a crash per kilometer of road. This could be explained by the shift from public transport to private car ownership with low vehicle occupancy. In both models, the parameter VI/RL showed a consistent negative estimate on annual road fatality figures. Road length as presented in VI/RL is often termed as an exposure component. The trend in the data shows a rise in the number of annual road deaths with a corresponding increase in total road length in the country as shown in Figure 3.

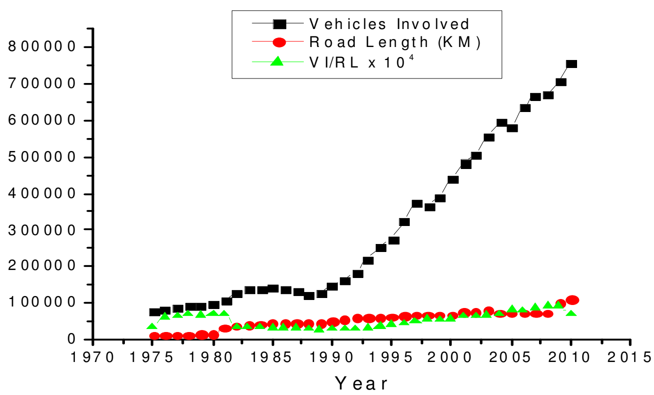

Figure 3 Trend of VI & RL

Mobile cellular subscription among the populace is shown to have a statistically significant effect on road accidents and road death in the second model. This is in agreement with the results of other studies for example; Fowles et.al [42] found that an increase in mobile phone usage leads to increase in accident rate in trucks, similar finding on motor vehicle accident was reported by [43, 44]. Equally, Wilki and Zlatoper [45] concluded on the existence of non-linear statistically significant impact of mobile phone subscriptions on road fatality rate. The significance of mobile cellular subscription in this study could be an indication of widespread usage of mobile phones by road users although this claim needs to be further investigated.

4.1 Models and Model Fit

To establish the significance of the selected predictor variables, two (2) models were fitted as shown in Table 2. Generalized Estimation Equations were used with Negative binomial and Poisson Log-linear distributions. The final models are in the form given in Equation 8;

$$\mu_n = \exp(\beta_o).\exp\left(\sum_{j=1}^{k}\beta_j X_j\right) \tag{8}$$

Where;

μ_n = Estimated fatality for any given year 'n'

β_o and β_j are regression coefficients of X_js

X_j = Model variables

Using Wald χ^2, the significance of each of the model variables using was established. The Wald χ^2 is similar to t-statistics used in regular regression analysis as elaborated by [46].

$R^2_{\alpha max}$ was used to account for how well the variance of the data is explained in a relative sense. Values of 0.939 and 0.457 were recorded for the first and second models respectively. Miaou [47] suggested the use of negative binomial overdispersion parameter in computing the $R^2_{\alpha max}$ thus;

$$R^2_{\alpha\max} = 1 - \frac{\alpha}{\alpha_{\max}} \tag{9}$$

Where;

α = estimated over-dispersion parameter for the selected model

α_{max} = estimated over-dispersion parameter for intercept only model

Quasi-likelihood (QIC) method under independent model criterion was used to compare the model fit. Table 2 shows the QIC and QICC values of the models. The QIC is a modification of the AIC (Akaike's Information Criterion) with proper adjustment given to the penalty term to suit the operational properties in the GEE. The QIC is reported to work well in the selection of working correlation matrix and variable selection [48, 49].

Table 2 Model Parameters

Variable (Annual Count)		Negative Binomial			Poisson Log-Linear		
Dependent Model One (P-VI/RL)	Independent	β	S.E	P-value	β	S.E	P-value
	Constant	7.173	0.0449	<0.001	7.205	0.0492	<0.001
	Population	0.066	0.0023	<0.001	0.066	0.0028	<0.001
	VI/RL	-0.029	0.0069	<0.001	-0.034	0.0075	<0.001
Observations (N)		36			36		
QIC (Quasi-Likelihood _independent model)		40.5700			1507.7950		
QICC (Corrected Quasi-Likelihood _independent model)		41.9440			1338.5980		
R^2_{α}		0.939			N/A		

Model two (VI/RL-Mc)							
	Constant	8.532	0.1218	<0.001	8.458	0.1228	<0.001
	VI/RL	-0.059	0.0286	0.040	-0.037	0.0298	0.215
	Mobile cells/100people	0.009	0.0018	<0.001	0.007	0.0017	<0.001
Observations (N)		36			36		
QIC (Quasi-Likelihood _independent model)		41.3490			11196.9870		
QICC (Corrected Quasi-Likelihood _independent model)		42.3810			9820.1610		
R^2_α		0.457			N/A		

4.2. Models Validation and Comparison

The models selected are presented in Equations 10 and 11. Model outputs are compared with the actual values recorded from 2011-2015 and results from some existing models as presented in Table 3. The intercept in Model VI/RL-Mc showed great influence and therefore, only model P-VI/RL was considered in the comparison.

For the year 2020, population is projected at 1.5% based on the current trend reported by the World Bank. Road length was projected using the annual growth rate of 1%. While 6.2% and 2.5% were used for vehicle Involved in crashes and Mobile Cellular Subscription per 100 people respectively.

The average yearly changes in the number of death for the actual values recorded in the period 2005 - 2015 is at about 0.6%. Comparatively, about 2.4% was recorded by the models formulated in this study. This shows the performance of the models when compared to about 2% shown by previous Smeed's Law model reported by[39] and 3% by the ARIMA model reported by [23].

The P-VI/RL and VI/RL-Mc models from this study are given as follows;

Model P-VI/RL, GEE-NBR:

$$Fat_n = 1303 \times e^{(66\times10^{-3} P_n - 29\times10^{-3} K_n)} \tag{10}$$

Model VI/RL-Mc, GEE-NB:

$$Fat_n = 5074 \times e^{(9\times10^{-3} Mc_n - 59\times10^{-3} K_n)} \tag{11}$$

Where: Fat_n = Number of fatality in a given year 'n'

Pn = Population $x10^{-6}$ projected for the year 'n'

K_n = VI/RL projected for the year 'n'

M_n = Mobile cell subscription/100people projected for the year 'n'

Table 3 Models comparison with actual values and some existing models

Model/Source	Form	Prediction					
		2011	2012	2013	2014	2015	2020
Recorded	**Actual**	**6877**	**6917**	**6915**	**6674**	**6706**	**N/A**
[39]	Smeed's Law	7397	7571	7750	7918	8115	9172
[40]	GLM-NBR	>20000 for the Whole Period					
[23]	ARIMA	N/A	N/A	8000	8300	8760	10716
Model P-VI/RL (This Study)	**GEE-NBR**	**7252**	**7304**	**7438**	**7669**	**8178**	**9342**

5. CONCLUSION

The Malaysian road safety record is not very encouraging compared to the level of investment by the government in the area. Considering global trends in road safety target settings and the result from this study, it could be a good call if other goal setting approaches different from empirical methods are explored in setting road safety targets.

There could be some gain if rural public transport is improved and car sharing is encouraged among commuters. This will reduce the fleet size on the roads thereby improving the accident and fatality rates.

Exposure variables that are shown to be significant in fatality prediction models presented in this study can be targetted in the plans towards achieving aspirational road safety targets. These variables are; Population, mobile cellular subscription, road length, and vehicle Involved in crashes.

Based on available records, this is the first time Mobile subscription is considered in fatality prediction modeling in Malaysia. The significance of mobile cellular subscription could be an indication of mobile phone usage by road users. The study shows that the risk of recording a fatality increases with mobile phone subscriptions. Therefore, significant measures should be taken in encouraging proper usage of mobile phones among drivers. This may perhaps be achieved through law enforcement and the use of modern technology.

The result of this study shows that GEE can be used to predict fatality for the short-term provided that overdispersion is accounted for. Using 2010 as the base year, the models from this study predicted total fatality of 8178 for the year 2015 compared to 8760 predicted for the same year by the ARIMA model of [23], both results are compared against the actually recorded fatality of 6706 for the year 2015. Equally, for the year 2020, the models from this study predicted total fatality of 9342 compared to 9172 and 10716 by Karim et.al and the ARIMA model by Rohayu et.al. The output from all the three models is closer to the actually recorded fatality figures when the 30% reduction target set by the Malaysian government is applied. Therefore, this model may be used as an additional tool for preliminary evaluation of the fatality reduction policy.

REFERENCES

[1] Who, World Health Organization global status report on road safety 2013: supporting a decade of action: World Health Organization, 2013.

[2] W. H. O. Who., Global Status Report on Road Safety 2015, 20 Avenue Appia, 1211 Geneva 27, Switzerland2015.

[3] G. Al-Haji, Road Safety Development Index: Theory, Philosophy and Practice, 2007.

[4] I. f. H. Metrics and Evaluation, The global burden of disease: generating evidence, guiding policy, ed: IHME Seattle^ eWA WA, 2013.

[5] W. H. Organization, Global health risks: mortality and burden of disease attributable to selected major risks: World Health Organization, 2009.

[6] World-Bank, Doing Business 2016: Measuring Regulatory Quality and Efficiency, ed: World Bank Washington, DC, 2016.

[7] R. J. Smeed, Some Statistical Aspects of Road Safety Research, Journal of the Royal Statistical Society. Series A (General), vol. 112, pp. 1-34, 1949.

[8] M. Sivak, Society's aggression level as a predictor of traffic fatality rate, Journal of Safety Research, vol. 14, pp. 93-99, 1983.

[9] R. Fieldwick and R. J. Brown, Effect of speed limits on road casualties, Traffic engineering and control, vol. 28, pp. 635-640, 1987.

[10] S. Hakim, D. Shefer, A. Hakkert, and I. Hocherman, A critical review of macro models for road accidents, Accident Analysis & Prevention, vol. 23, pp. 379-400, 1991.

[11] G. Marsden and P. Bonsall, Performance targets in transport policy, Transport policy, vol. 13, pp. 191-203, 2006.

[12] Y.-J. Kweon, Data-driven reduction targets for a highway safety plan, Transport policy, vol. 17, pp. 230-239, 2010.

[13] L.-Y. Chang and F. Mannering, Analysis of injury severity and vehicle occupancy in truck-and non-truck-involved accidents, Accident Analysis & Prevention, vol. 31, pp. 579-592, 1999.

[14] M. Bedard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, The independent contribution of driver, crash, and vehicle characteristics to driver fatalities, Accident Analysis & Prevention, vol. 34, pp. 717-727, 2002.

[15] K. E. MacLeod, J. B. Griswold, L. S. Arnold, and D. R. Ragland, Factors associated with hit-and-run pedestrian fatalities and driver identification, Accident Analysis & Prevention, vol. 45, pp. 366-372, 2012.

[16] M. F. Ballesteros, P. C. Dischinger, and P. Langenberg, Pedestrian injuries and vehicle type in Maryland, 1995–1999, Accident Analysis & Prevention, vol. 36, pp. 73-81, 2004.

[17] D. Plurad, D. Demetriades, G. Gruzinski, C. Preston, L. Chan, D. Gaspard, et al., Pedestrian injuries: the association of alcohol consumption with the type and severity of injuries and outcomes, Journal of the American College of Surgeons, vol. 202, pp. 919-927, 2006.

[18] A. V. Moudon, L. Lin, J. Jiao, P. Hurvitz, and P. Reeves, The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington, Accident Analysis & Prevention, vol. 43, pp. 11-24, 2011.

[19] H. Z. Dovom, M. S. M. Z. Dovom, and N. Nadimi, An Analysis of Pedestrian Fatal Accident Severity Using a Binary Logistic Regression Model, Institute of Transportation Engineers. ITE Journal, vol. 82, p. 38, 2012.

[20] V. Bewick, L. Cheek, and J. Ball, Statistics review 14: Logistic regression, Crit Care, vol. 9, pp. 112-118, 2005.

[21] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time series analysis: forecasting and control: John Wiley & Sons, 2015.

[22] K. Adu-Poku, R. Avuglah, and E. Harris, Modeling Road Traffic Fatality Cases in Ghana.

[23] S. Rohayu, S. Sharifah Allyana, M. Jamilah, and S. Wong, Predicting Malaysian road fatalities for year 2020, 2012.

[24] Y.-Y. Pang, X.-J. Zhang, Z.-B. Tu, M.-J. Cui, and Y. Gu, [Autoregressive integrated moving average model in predicting road traffic injury in China], Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi, vol. 34, pp. 736-739, 2013.

[25] M. A. Quddus, Time series count data models: an empirical application to traffic accidents, Accident Analysis & Prevention, vol. 40, pp. 1732-1741, 2008.

[26] S. C. Joshua and N. J. Garber, Estimating truck accident rate and involvements using linear and Poisson regression models, Transportation planning and technology, vol. 15, pp. 41-58, 1990.

[27] S.-P. Miaou, P. S. Hu, T. Wright, A. K. Rathi, and S. C. Davis, Relationship between truck accidents and highway geometric design: a Poisson regression approach, Transportation Research Record, 1992.

[28] V. Shankar, F. Mannering, and W. Barfield, Effect of roadway geometrics and environmental factors on rural freeway accident frequencies, Accident Analysis & Prevention, vol. 27, pp. 371-389, 1995.

[29] G. S. Tulu, S. Washington, M. M. Haque, and M. J. King, Investigation of pedestrian crashes on two-way two-lane rural roads in Ethiopia, Accident Analysis & Prevention, vol. 78, pp. 118-126, 2015.

[30] S.-P. Miaou, The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, Accident Analysis & Prevention, vol. 26, pp. 471-482, 1994.

[31] D. Lord, S. P. Washington, and J. N. Ivan, Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory, Accident Analysis & Prevention, vol. 37, pp. 35-46, 2005.

[32] 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, pp. 291-305, 2010.

[33] P. C. Anastasopoulos and F. L. Mannering, A note on modeling vehicle accident frequencies with random-parameters count models, Accident Analysis & Prevention, vol. 41, pp. 153-159, 2009.

[34] P. K. Andersen, 1. Generalized estimating equations. James W. Hardin and Joseph M. Hilbe, Chapman and Hall/CRC, Boca Raton, 2003. No. of pages: xiii+ 222 pp. Price: $79.95. ISBN 1-58488-307-3, Statistics in Medicine, vol. 23, pp. 2479-2480, 2004.

[35] G. M. Fitzmaurice, N. M. Laird, and A. G. Rotnitzky, Regression models for discrete longitudinal responses, Statistical Science, pp. 284-299, 1993.

[36] K.-Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika, vol. 73, pp. 13-22, 1986.

[37] R. A. Memon, G. Khaskheli, and A. S. Qureshi, Operating speed models for two-lane rural roads in Pakistan, Canadian Journal of Civil Engineering, vol. 35, pp. 443-453, 2008.

[38] The-World-Bank, "World Development Indicators," T. W. Bank, Ed., ed, 2016.

[39] M. R. Karim, A macro analysis of road accident trends in Malaysia, Journal of the Eastern Asia Society for Transportation Studies, vol. 1, pp. 941-950, 1995.

[40] R. Sohadi and R. Umar, Model kematian jalan raya di Malaysia: unjuran tahun 2000, Pertanika Journal of Science & Technology, vol. 6, pp. 107-119, 1998.

[41] L. Persia, R. Gigli, and D. S. Usami, Smeed's law and expected road fatality reduction: an assessment of the Italian case, Journal of safety research, vol. 55, pp. 121-133, 2015.

[42] R. Fowles, P. D. Loeb, and W. Clarke, The cell phone effect on truck accidents: A specification error approach, Transportation research part E: logistics and transportation review, vol. 50, pp. 18-28, 2013.

[43] E. Muehlegger and D. Shoag, Cell phones and motor vehicle fatalities, Procedia engineering, vol. 78, pp. 173-177, 2014.

[44] D. Sun and A. Jia, Impacts of cell phone use on driving safety and drivers’ perception of risk, Journal of Modern Transportation, vol. 24, pp. 145-152, 2016.

[45] A. M. Welki and T. J. Zlatoper, The effect of cell phones on international motor vehicle fatality rates: A panel-data analysis, Transportation research part E: logistics and transportation review, vol. 64, pp. 103-109, 2014.

[46] A. Agresti and M. Kateri, Categorical data analysis: Springer, 2011.

[47] S.-P. Miaou, Measuring the goodness-of-fit of accident prediction models, 1996.

[48] W. Pan, Akaike's information criterion in generalized estimating equations, Biometrics, vol. 57, pp. 120-125, 2001.

[49] J. W. Hardin and J. M. Hilbe, Generalized estimating equations: Wiley Online Library, 2003.

[50] Jasvinder Singh, Mahipal Singh, Anil Baliram Ghubade and Manjinder Singh Analytical Hierarchy Process for Road Accident of Motorcycle in India: A Case Study. International Journal of Mechanical Engineering and Technology, 8(7), 2017, pp. 1348–1356.

[51] Gezim Hoxha, Ahmet Shala, Ramë Likaj, Pedestrian Crash Model for Vehicle Speed Calculation at Road Accident. International Journal of Civil Engineering and Technology, 8(9), 2017, pp. 1093–1099.

[52] B.Naga Kiran, Dr. N. Kumara Swamy and Dr. C. Sashidhar, Prediction o Road Accident Modelling For Indian National Highways, International Journal of Civil Engineering and Technology, 8(1), 2017, pp.789-802.