

Modelling Road Traffic Accident Costs in Developing Countries

Bello Kabiru^{1,*}, Abdulfatai Adinoyi Murana¹ & Adekunle Taiwo Olowosulu¹

*Corresponding author: kabirulado@gmail.com

¹Department of Civil Engineering, Ahmadu Bello University Zaria, P.M.B. 1045 Samaru Zaria, Nigeria

ABSTRACT

Researchers have been yearning for a mathematical model of costing road traffic accidents due to the shortcomings of the existing methods. This research is aimed at developing the mathematical model for road traffic accident costs in developing countries. To achieve the aim, basic road traffic accident cost determinants (death rate, interest rate, Gross National Income, life expectancy, average age at first marriage, literacy and population) were sourced from the websites of World Bank, World Health Organization (WHO), International Statistical Institute (ISI), International Monetary Fund (IMF), United Nations (UN), Central Intelligence Agency (CIA), Google, Yahoo and Ask. Then, some of the basic accident cost determinants were used to drive Gross National Income loss, cost of casualty by life insurance approach, amount of road safety and accident severity number using a developed program written in FORTRAN 77 language. With the derived parameters and the remaining basic parameters, goodness of fit was carried out using Easyfit 5.5 professional to choose the relevant explanatory variables for the response variable and to obtain the appropriate regression analysis for developing the model. Log – logistic regression analysis was used to develop the model which expresses transformed Gross National Income loss (logit) in terms of life expectancy (years), amount of road safety and accident severity number (with no unit). When tested, the model was not able to predict the Gross National Income loss of some developing countries through logit transformation formula of its response variable and therefore a complementary model was developed using Graph 4.3.0.384 to obtain unlimited logit transformation of the response variable. The complementary model gives the Gross National Income loss in dollars per year and it was checked against the existing methods for accuracy. Thereby, it was found out that the model is only statistically the same as the other methods of costing road traffic accident at no risk. However, the model is statistically the same as the human capital approach of costing road traffic accident even at an extreme risk of 42.5%. Therefore, the model could be used for decision making to invest in road safety of developing countries.

© 2021 Malaysian Institute of Road Safety Research (MIROS). All Rights Reserved.

ARTICLE INFO

Article History:

Received 15 Mac 2021

Received in revised form

17 Apr 2021

Accepted

28 Apr 2021

Available online

01 May 2021

Keywords:

Mathematical Model

Road Traffic Accident Costs

Developing Countries

Gross National Income Loss

Extreme Risk

Road Safety

1. Introduction

Each day, almost 3,700 people are killed globally in road traffic crashes and more people now die in the crashes than from HIV/AIDS (Centers for Disease Control and Prevention [CDC], 2019). As the number of motor vehicles and vehicle – miles of travel increase throughout the world, the exposure of the population to traffic crashes also increases (Garber & Hoel, 2009). The annual global number of road traffic accident fatalities has swelled by around 100,000 in just three years (Sadik-Khan & Solomonow, 2020). The number of road traffic deaths was estimated at 1.25 million annually (WHO, 2015).

The number of annual road traffic deaths has reached 1.35 million and road accidents kill someone every 24 seconds (WHO, 2018). According to International (2014) every 30 seconds someone, somewhere, dies in a road crash and ten are seriously injured. Odero et al. (1997) stated that nearly three-quarters of road deaths occur in developing countries and men comprise a mean 80% of casualties.

Developing countries tend to have some characteristics in common due to their histories or geographies (“Developing country,” 2021). The distribution of accident rates in developing countries is peculiar and distinct from developed countries in which there are different road conditions, traffic traits, driver characteristics and government policies. This is argued by horizontal grouped bar charts in Figures 1 and 2 for the 2011 and 2015 WHO road traffic accident publications respectively.

The figures show high number of developing countries had high accident rates while low number of developed countries had high accident rates. However, low number of developing countries had low accident rates while high number of developed countries had low accident rates.

Road traffic crashes cost countries approximately 3% of their gross national product and this figure rises to 5% in some low-and middle-income countries (Iqbal & Zafar, 2017). According to Odero et al. (1997) applying the human capital approach Kenya lost \$14 million in 1984 by road traffic crashes which accounted

approximately to 1.6% of its Gross National Product. The annual losses in developing countries occasioned by Road Traffic Accidents exceed the annual development aid loans received by these countries (World Bank, 2010). Each year, according to the latest available cost estimate (1998), road traffic injuries cost \$518 billion USD worldwide and \$65 billion USD in low – and middle – income countries, which exceeds the total amount that these countries receive in the development assistance.

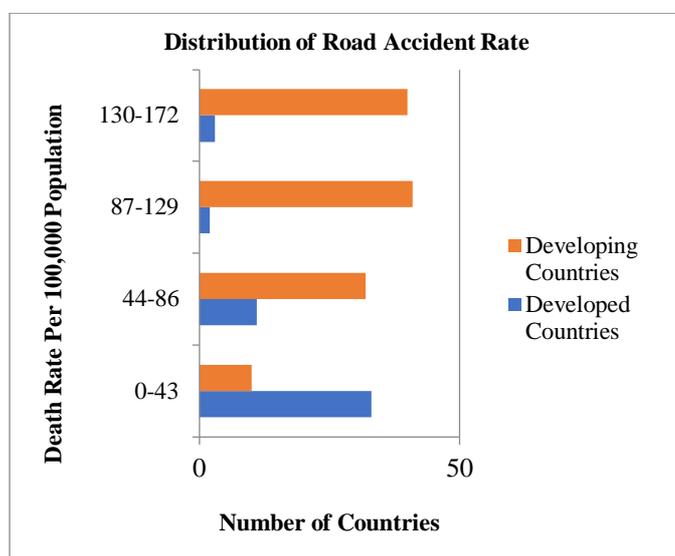


Figure 1: Death rates per 100,000 data (WHO, 2011)

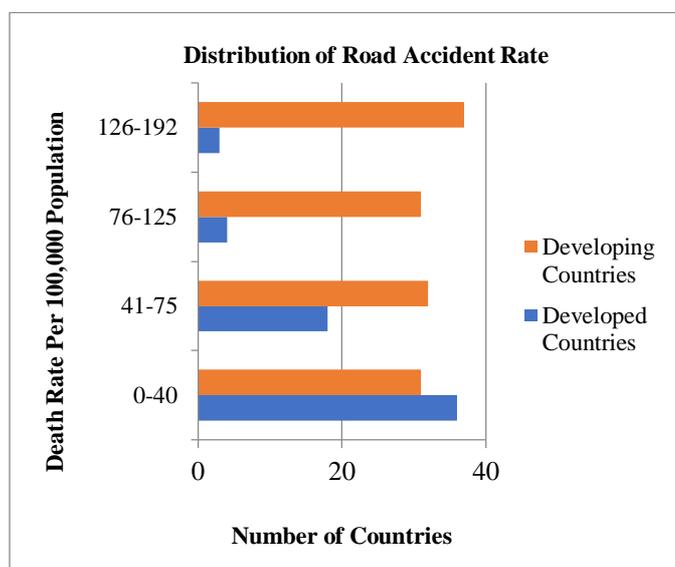


Figure 2: Death rates per 100,000 data (WHO, 2015)

Many researches modelled road traffic accident such as: (a) Fletcher et al. (2006) road accident rate model; (b) Benson (1996) and Garba (2014) road accident cost models; (c) Hejazi et al. (2012) road traffic accident losses; (d) Abdelwahab and Abdel-Arty (2001, 2004) road traffic accident injury levels models; (e) Julius et al. (2015) road traffic accident number of injuries model; (f) Abojaradeh (2013), Abojaradeh et al. (2013) and Abojaradeh et al. (2017) road traffic accident number; and so on.

Six methods of assessing monetary value of a road traffic accident victim exist which have been in use for many years in the literature. Hills and Jones-Lee (1983) as well as Jacobs et al. (2000) discussed the six methods. The methods are gross output/human capital method, net output method, life - insurance method, court

award method, implicit public sector valuation method and willingness-to-pay method.

Mathematical model for costing road traffic accidents is not yet found after the researches of Garba (2014) which used gross output approach, Benson (1996) which used net gross output approach, and Hejazi et al. (2012) which used life insurance approach and poison regression analysis. Although there is no real substitute, in individual countries, to carry out a detailed appraisal of national accident costs, expressing accident costs as a percentage of Gross National Income (GNI) provides a useful, albeit crude; approach to cost accidents, particularly on a global or regional basis (Richard & Bent, 2004). So, there is no mathematical model for expressing road traffic accident costs as a percentage of Gross National Income (GNI). Therefore, a mathematical model of costing road traffic accidents in terms of Gross National Income (GNI) will be developed in this study.

The objectives to develop the model are:

- Determine the developing countries Gross National Income loss, cost of casualty and accident severity number.
- Develop a mathematical model for predicting Gross National Income loss by road traffic accidents for the developing countries.
- Test the model for accuracy by comparing its prediction values and prediction values of the existing six methods of costing road traffic accident.

The countries in this study were classified as developing based on their Gross National Income. The loss of Gross National Income of developing countries was probabilistic instead of historic due to its unavailability.

2. Method

2.1. Data Source

Data were sourced in this study to give some independent variables and input for computation of some other explanatory variables from international organizations publications databases.

- 2015 - List of developing countries by the ISI.
- 2013 - Gross National Income of countries by the World Bank.
- 2013 - Literacy in percent of countries by the CIA.
- 2013 - Road accident death rate by country publication of the WHO.
- 2013 - Inflation rate of countries by the CIA.
- 2015 - Human life expectancy by country publication of the WHO.
- 2017 - Average age of first marriage by Wikipedia.
- 2013 - Population of countries by UN.

2.1.1. Sampling

According to the ISI, there are 137 developing countries in 2015. To save time, money and resources, a sample was used, instead of the entire population of the developing countries. However, a sample size of approximately 120 developing countries which is close to the entire population of 137 developing countries was obtained using SurveyMonkey's sample size calculator at 90% confidence level and 3% margin of error. The 120 developing countries were selected with replacement by simple random sampling.

2.1.2. Finding Gross National Income loss

The real values of Gross National Income loss by road traffic accidents for countries of the world are not available or published by anybody. Therefore, the values were derived from the World Bank and WHO publications (i.e. 3.0% of GNP and 5.0% of GNP). They were simulation of how road traffic accidents and costs are randomly assigned naturally. Then, they were made empirical by iteration in this study.

$$GNI\ loss = (a/100)b \tag{1}$$

where:

- GNI loss = Gross National Income loss value
- a = Percentage value of Gross National Income loss
- b = Gross National Income

Gross National Product losses were assigned to the 120 selected countries from 3.0% to 5.0 % (based on WHO publication) at interval of 0.1% also by simple random sampling.

2.2. Data Analysis

2.2.1. Gross National Income loss

The data of Gross National Income loss was automatically fitted to a large number of distributions by using Easyfit software and thereby the best fitting distribution model was found as log – logistic distribution. The data was transformed in order to use normal multiple regression analysis by finding its logit transformation.

$$Y = \text{logit}(y) = \ln(y/1 + y) \tag{2}$$

$$y = \text{logit}(Y) = (e^Y / 1 - e^Y) \tag{3}$$

Note that Equations 2 and 3 were derived from Equations 4 and 5 for logit transformation by changing their signs.

$$l = \text{logit}(p) = \ln(P/1 - P) \tag{4}$$

$$p = \text{logit}(l) = (e^l / 1 + e^l) \tag{5}$$

The signs were changed because:

- i. Logistic regression using Equations 4 and 5 requires a continuous response variable with values between 0 and 1. Some of this study values are more than 1 and therefore logit transformation results in natural logarithm of a negative number which calculator displays ‘Math ERROR’.
- ii. Goodness of fit methodology suggests that Log – Logistic regression is the appropriate approach for developing the model in this study and therefore the need to make Equations 4 and 5 to work for the regression analysis arose. Changing the signs of Equations 4 and 5 to give Equations 2 and 3 solved the problem.

2.2.2. Cost of casualty and accident severity number

Calculations were done based on the obtained road traffic accident cost data. Cost of casualty was obtained based on life insurance in Table 1 (using the difference between average age at first marriage and life expectancy), Equation 6 to account for time value of money and then Equation 7 to calculate the cost of casualty.

Table 1: Average global cost of life insurance

| Insurance term (years) | 10 | 15 | 20 | 30 | Renewal |
|------------------------|-----|-----|-----|-------|---------|
| Annual cost (\$) | 569 | 705 | 905 | 1,528 | 565 |

Source: www.valuepenguin.com/average-cost-life-insurance

$$P(\text{life insurance cost}) = A \left[\frac{1-(1+i)^n}{i} \right] \tag{6}$$

$$x_2 = [(u_1 \times f_1) + (u_2 \times f_2) + (u_3 \times f_3)] \tag{7}$$

where:

- x_2 = Cost of casualty
- u_1 = Cost of death

- u_2 = Cost of injury
- u_3 = Cost of property damage
- f_1 = Number of dead people
- f_2 = Number of injured people
- f_3 = Number of property damage only accident people

Accident severity number with no unit was calculated based on Garber (2009) weighing scale (fatality = 12, personal injury = 3 and property damage only accident = 1) and Equation 8.

$$x_3 = [(s_1 \times f_1) + (s_2 \times f_2) + (s_3 \times f_3)](10^{-6}) \tag{8}$$

where:

- x_3 = Accident severity number
- s_1 = Fatality weight
- s_2 = Personal injury weight
- s_3 = Property damage only accident weight

Computer programs were written in FORTRAN 77 language to facilitate the work, avoid mistakes in calculation and reduce time of doing the work.

2.2.3. Developing road traffic accident cost model

Linear least squares regression method was used with Gross National Income loss transformed as the response variable (y) and the following explanatory variables:

- i. Population (x_1) as the number of people in a developing country. Tuvalu has the minimum value of population, 10,000 people and Indonesia has the maximum value of population, 234,181,400 people.
- ii. Cost of casualty (x_2) in billion dollars. Indonesia has the highest cost of casualty of \$8.07 while Grenada, Palau, Samoa, Seychelles and St. Vincent have the lowest cost of casualty of less than \$10m.
- iii. Accident severity number (x_3) which has no unit because it was derived from addition of basic quantities (Gupta, 2009). Indonesia has the highest accident severity number of 1.04 while Grenada, Palau, Samoa, Seychelles and St. Vincent have the lowest accident severity number of less than 10,000.
- iv. Inflation rate (x_4) in percent. Bulgaria and Argentina have the highest deflation and inflation rates of -1.4% and 36.4% in that order while Macedonia and Fiji have the lowest deflation and inflation rates of -0.1% and 0.1% respectively.
- v. Life expectancy (x_5) in years. Sierra Leone has the minimum value of the human life expectancy of 46 years while Lebanon has the maximum value of the human life expectancy of 80 years.
- vi. Literacy (x_6) in percent. Azerbaijan has the highest literacy value of 99.8% while South Sudan has the lowest literacy value of 27.0%.

2.2.4. Model testing

In testing the model, relevant information of developing countries corresponding to known values of road accidents were obtained. Then some determinants of the model were found using Equations 7 and 8.

The corresponding determinants of the costs were used to test the model by comparing its predictions (observed values) to the existing methods predictions (expected values).

3. Results

3.1. Goodness of Fit Results

3.1.1. Derived response variable fit distributions

The fit distributions of the Gross National Income loss in Table 2 shows Log - Normal distribution ranked first and second in the Kolmogorov Smirnov test.

Table 2: Fit distributions of the derived response variable (Kolmogorov Smirnov)

| No. | Distribution | Kolmogorov Smirnov | |
|-----|--------------------|--------------------|------|
| | | Statistic | Rank |
| 1 | Log - Normal | 0.05362 | 1 |
| 2 | Log - Normal | 0.05362 | 2 |
| 3 | Pareto 2 | 0.05725 | 3 |
| 4 | Gen. Pareto | 0.10418 | 15 |
| 5 | Gen. Extreme Value | 0.11722 | 19 |
| 6 | Dagum | 0.06577 | 8 |
| 7 | Log - Logistic | 0.06226 | 4 |
| 8 | Inv. Gaussian | 0.10382 | 14 |
| 9 | Pareto 2 | 0.05725 | 3 |

In Table 3, Gen. Pareto was ranked first in the Anderson Darling test for fit distributions of the Gross National Income loss.

Table 3: Fit distributions of the derived response variable (Anderson Darling)

| No. | Distribution | Anderson Darling | |
|-----|--------------------|------------------|------|
| | | Statistic | Rank |
| 1 | Log - Normal | 2.5593 | 13 |
| 2 | Log - Normal | 2.5593 | 14 |
| 3 | Pareto 2 | 2.3573 | 10 |
| 4 | Gen. Pareto | 1.6413 | 1 |
| 5 | Gen. Extreme Value | 2.137 | 2 |
| 6 | Dagum | 2.3167 | 3 |
| 7 | Log - Logistic | 2.4483 | 11 |
| 8 | Inv. Gaussian | 15.615 | 19 |
| 9 | Pareto 2 | 2.3573 | 10 |

Chi - Squared test ranked Log - Logistic first for fit distributions of the Gross National Income loss as shown in Table 4.

Table 4: Fit distributions of the derived response variable (Chi-Squared)

| No. | Distribution | Chi - Squared | |
|-----|--------------------|---------------|------|
| | | Statistic | Rank |
| 1 | Log - Normal | 7.9338 | 6 |
| 2 | Log - Normal | 7.9338 | 7 |
| 3 | Pareto 2 | 7.1134 | 3 |
| 4 | Gen. Pareto | 10.571 | 15 |
| 5 | Gen. Extreme Value | 18.488 | 21 |
| 6 | Dagum | 8.2762 | 10 |
| 7 | Log - Logistic | 4.492 | 1 |
| 8 | Inv. Gaussian | 7.0562 | 2 |
| 9 | Pareto 2 | 7.1134 | 3 |

However, combining the three tests results rank Pareto 2 and Log - Logistic distributions first because they have the lowest ranking score of 16 points as shown in Table 5.

Table 5: Combined test results

| No. | Distribution | Ranking Score |
|-----|--------------------|---------------|
| 1 | Log - Normal | 20 |
| 2 | Log - Normal | 23 |
| 3 | Pareto 2 | 16 |
| 4 | Gen. Pareto | 31 |
| 5 | Gen. Extreme Value | 42 |
| 6 | Dagum | 21 |
| 7 | Log - Logistic | 16 |
| 8 | Inv. Gaussian | 35 |
| 9 | Pareto 2 | 16 |

Therefore, based on the fit distributions of the Gross National Income loss results, the model can be developed using:

- i. Log - Normal regression analysis
- ii. Pareto 2 regression analysis
- iii. Log - Logistic regression analysis

3.1.2. Explanatory variables fit distributions

The result of fit distributions for the relevant explanatory variables is shown in Table 6-8.

Table 6: Fit distributions of the explanatory variables (Kolmogorov Smirnov test)

| No. | Explanatory Random Variable | | Kolmogorov Smirnov Test Best Distribution Model |
|-----|-----------------------------|----------|---|
| | Name | Notation | |
| 1 | Population | x_1 | Log - Pearson 3 |
| 2 | Cost of casualty | x_2 | Log - Logistic |
| 3 | Accident severity number | x_3 | Gen. Pareto |
| 4 | Inflation rate | x_4 | Burr (4p) |
| 5 | Life expectancy | x_5 | Gen. Extreme Value |
| 6 | Literacy | x_6 | Beta |

Table 7: Fit distributions of the explanatory variables (Anderson Darling test)

| No. | Explanatory Random Variable | | Anderson Darling Best Distribution Model |
|-----|-----------------------------|----------|--|
| | Name | Notation | |
| 1 | Population | x_1 | Weibull |
| 2 | Cost of casualty | x_2 | Gen. Pareto |
| 3 | Accident severity number | x_3 | Gen. Pareto |
| 4 | Inflation rate | x_4 | Dagum (4p) |
| 5 | Life expectancy | x_5 | Kumaraswamy |
| 6 | Literacy | x_6 | Johnson SB |

Table 8: Fit distributions of the explanatory variables (Chi-Squared test)

| No. | Explanatory Random Variable | | Chi - Squared Best Distribution Model |
|-----|-----------------------------|----------|---------------------------------------|
| | Name | Notation | |
| 1 | Population | x_1 | Fatigue Life |
| 2 | Cost of casualty | x_2 | Gen. Extreme Value |
| 3 | Accident severity number | x_3 | Gen. Pareto |
| 4 | Inflation rate | x_4 | Gamma (3p) |
| 5 | Life expectancy | x_5 | Triangular |
| 6 | Literacy | x_6 | Weibull |

The goodness of fit result summary is shown in Figure 3 in which it is clear that it is a simple methodology for building model in addition to the best subsets regression and the stepwise regression.

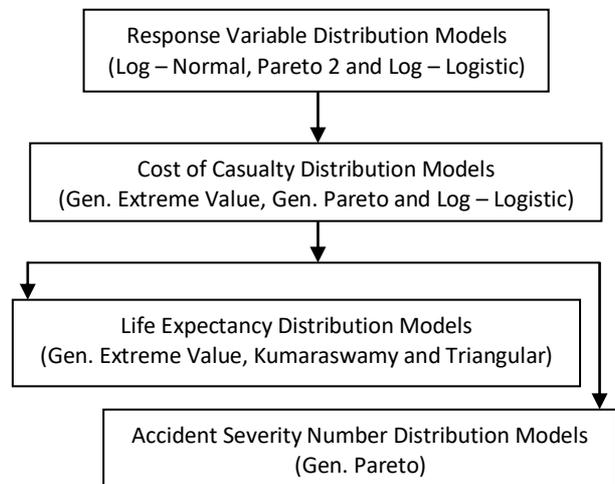


Figure 3: Goodness of fit result summary

3.2. Developing the Model Result

$$Y = 0.014x_5 - 0.329x_2 + 4.460x_3 - 2.160 \quad (9)$$

where:

Y = Transformation value of the Gross National Income loss in logit (unit of measurement)

x_5 = Life expectancy in years

x_2 = Cost of casualty in billion U.S. dollars

x_3 = Accident severity number with no unit

Gupta (2009) stated that derived quantities are those that may be expressed in terms of base or derived quantities by means of the mathematical symbols of multiplication and division only (no addition or subtraction or any other sign). Therefore, accident severity number has no unit because it was derived from Equation 8 which is an expression with addition sign.

3.2.1. Graphs of the model variables

The graph with regression of the log odds and life expectancy as dependent and independent variables respectively at 95% confidence level and 0.05 P -value is shown in Figure 4. It was found out that R^2 value was 4.9% and the P -value was 0.016. Therefore, life expectancy is statistically significant in the model as P -value = 0.016 < 0.05.

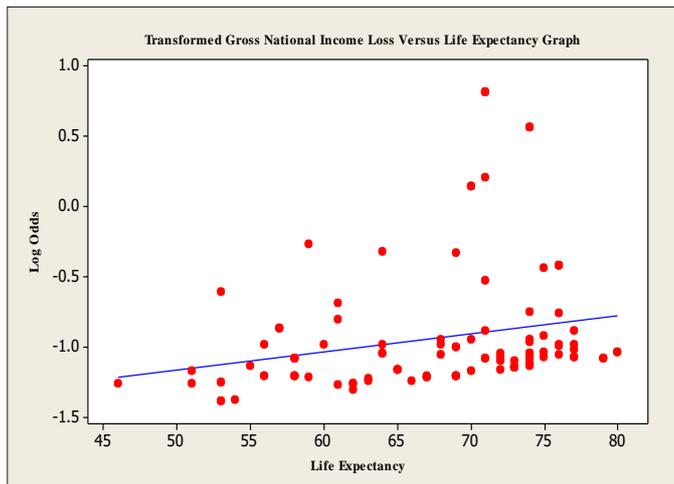


Figure 4: Graph of Y against x_5

The graph with regression of the log odds and cost of casualty as response and explanatory variables respectively at 95% confidence level and 0.05 P -value is shown in Figure 5. It was found out that R^2 value was 86.2% and the P -value was 0.000. Therefore, cost of casualty is statistically significant in the model as P -value = 0.000 < 0.05.

Figure 6 shows the graph with regression of the log odds and accident severity number as outcome and predictor variables respectively at 95% confidence level and 0.05 P -value. It was found out that R^2 value was 92.7% and the P -value was 0.000. Therefore, accident severity number is statistically significant in the model as P -value = 0.000 < 0.05.

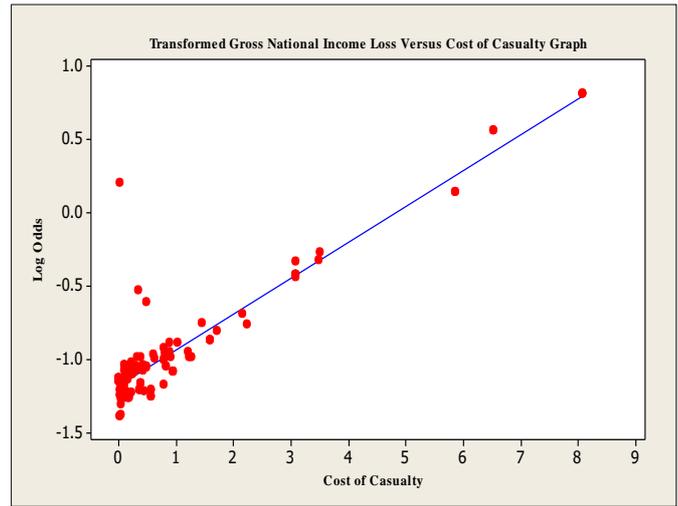


Figure 5: Graph of Y against x_2

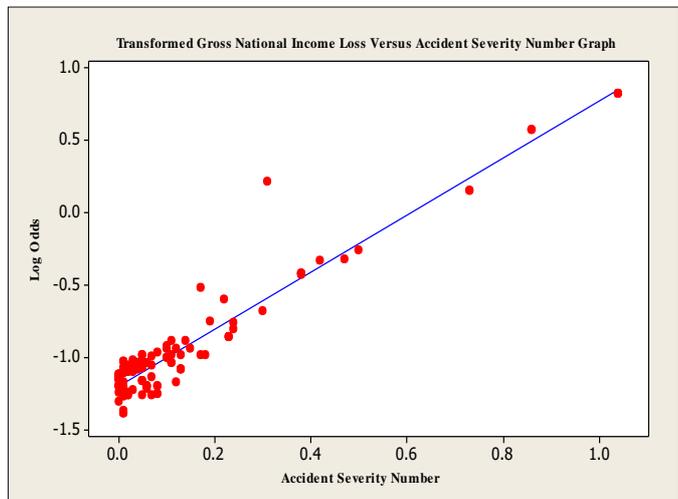


Figure 6: Graph of Y against x_3

3.2.2. Coefficients of the model variables

The odd in this study is the ratio of Gross National Income loss in dollars to the Gross National Income loss in dollars increased by one dollar.

The coefficient of life expectancy in the model is 0.014 which means that the natural logarithm of the odds increases by 0.014 for each additional one year. It agrees with the graph in Figure 4. Cost of casualty coefficient of -0.329 means that the natural logarithm of the odds increases by 0.329 for each deduction of one dollar. It contradicts the graph in Figure 5. Finding out that as cost of casualty increases the natural logarithm of the odds decreases is an indication of wrong sign.

Accident severity number coefficient of 4.460 means that the natural logarithm of the odds increases by 4.460 for each additional one unit. It agrees with the graph in Figure 6.

3.2.3. Correction for the wrong sign of cost of casualty in the model

Kennedy (2002) stated that an omitted explanatory variable may be the culprit for a wrong sign. Addition of amount of road safety did not address the problem but when the cost of casualty was replaced with the amount of road safety the problem was resolved.

$$x_7 = \frac{W.H.O \text{ Road Accidents Rank}}{10,000} \times \frac{Population}{100,000} \quad (10)$$

where:

x_7 = the amount of road safety in a developing country

Values of x_7 were calculated for each developing country using Equation 10, WHO (2018) road traffic accident ranking and the UN (2013) population of countries. The corrected model is shown in Equation 11.

$$Y = 0.013x_5 - 0.019x_7 + 2.280x_3 - 2.060 \quad (11)$$

It was found out that R^2 value was 81.5% and the P -value was 0.000. Therefore, amount of road safety is statistically significant in the model as P -value = 0.000 < 0.05. Its coefficient means that the natural logarithm of the odds decreases by 0.019 for each additional one unit of amount of road safety.

3.3. Calibrations

The original model in Equation 9 gave some negative prediction values and therefore, means of complementing its prediction process was developed graphically using Johansen (2007) Graph 4.3 Build 384 software as shown in Figure 7.

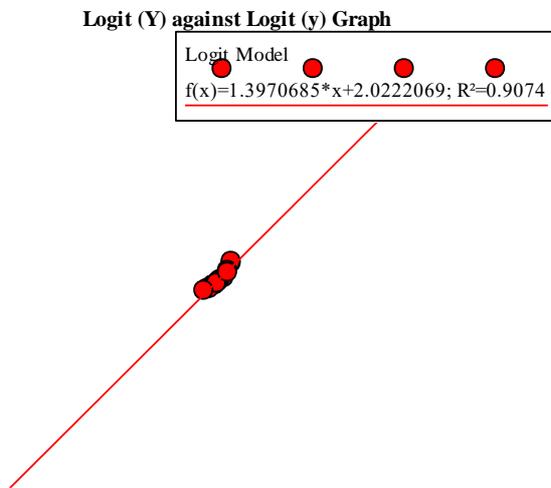


Figure 7: Log - Logistic graph

The graph results show coefficient of determination of 90.74% between the data and their function is

$$f(x) = 1.3970685x + 2.0222069 \quad (12)$$

where:

$f(x) = y$ = Gross National Income loss in U.S. dollars per year
 $x = Y$ = Transformed Gross National Income loss in logit.

Substituting $f(x)$ and x for y and Y respectively, Equation 12 becomes

$$y = 1.3970685Y + 2.0222069 \quad (13)$$

Equations 9 and 13 were used together as the final Log – Logistic regression analysis model and the prediction results are shown in Table 9.

Table 9: Complemented model prediction

| No. | Country | Y [Logit (y)] | y [Logit (Y)] | y (Existing Methods) |
|-----|--------------|------------------|------------------|-------------------------|
| 1 | Albania | -1.02 | 0.60 | 4.00 |
| 2 | Bangladesh | -0.75 | 0.98 | 0.24* |
| 3 | Brazil | 3.41 | 6.80 | 0.16 |
| 4 | Cameroon | -1.21 | 0.34 | 1.20 |
| 5 | Egypt | 4.75 | 8.68 | 1.30 |
| 6 | Egypt | 4.19 | 7.90 | 6.00 |
| 7 | Egypt | 4.12 | 7.80 | 0.97* |
| 8 | Egypt | 4.26 | 8.00 | 9.52 |
| 9 | Egypt | 2.27 | 5.21 | 0.39 |
| 10 | Ethiopia | -0.87 | 0.81 | 0.41 |
| 11 | Fiji | -1.18 | 0.38 | 0.19 |
| 12 | Ghana | -0.93 | 0.73 | 14.00 |
| 13 | India | 0.93 | 3.33 | 0.11 |
| 14 | Iran | 0.76 | 3.09 | 5.00 |
| 15 | Jordan | -0.92 | 0.74 | 0.41 |
| 16 | Kenya | -0.84 | 0.85 | 4.00 |
| 17 | Lebanon | -0.98 | 0.66 | 0.24 |
| 18 | Libya | -0.88 | 0.80 | 0.16 |
| 19 | Morocco | -0.52 | 1.30 | 1.20 |
| 20 | Morocco | -0.51 | 1.32 | 1.30 |
| 21 | Nigeria | 0.00 | 2.03 | 6.00 |
| 22 | Singapore | -1.17 | 0.39 | 0.97* |
| 23 | South Africa | 2.26 | 5.19 | 9.52* |
| 24 | Sudan | -0.87 | 0.81 | 0.39* |
| 25 | Sudan | 0.45 | 2.66 | 0.41* |
| 26 | Tunisia | -0.90 | 0.77 | 0.19 |
| 27 | Turkey | -0.61 | 1.18 | 14.00 |
| 28 | Uganda | -0.96 | 0.69 | 0.11 |
| 29 | Ukraine | -0.64 | 1.13 | 5.00 |
| 30 | Zimbabwe | -1.31 | 0.20 | 0.41 |

3.3.1. Hypothesis testing

Firstly, it was hypothesized that the observed values (Log - Logistic regression model values) and the expected values (Existing methods values in Table 4) are not the same and the test was carried out at different significance level as shown below.

H_0 : Observed values = Expected values

H_a : Observed values \neq Expected values

$n_1 = 30, v_1 = n_1 - 1 = 29$

$n_2 = 30, v_2 = n_2 - 1 = 29$

Using SPSS from Table 4:

$$s_1^2 = 137.564$$

$$s_2^2 = 7.535$$

where:

s_1^2 = Variance of existing methods values

s_2^2 = Variance of Log – Logistic values

$$F = s_1^2/s_2^2 = 18.257$$

From the F distribution table and calculator:

$$F_{0.10}(29,29) = 1.62$$

$$F_{0.05}(29,29) = 1.86$$

$$F_{0.025}(29,29) = 2.10$$

$$F_{0.01}(29,29) = 2.42$$

$$F_{0.001}(29,29) = 3.29$$

$$F_{0.00}(29,29) = 32.00$$

Calculated F is greater than critical value of F at all the significance levels except no risk as shown above. Therefore, the

Log – Logistic regression model is only statistically the same as the other methods of costing road traffic accident at no risk.

Secondly, it was hypothesized that the observed values (Log – Logistic regression model values) and the expected values (only human capital approach in Table 4 with asterisk) are not the same and the test was repeated as shown below.

H_0 : Observed values = Expected values

H_a : Observed values \neq Expected values

$n_1 = 30, v_1 = n_1 - 1 = 29$

$n_2 = 6, v_2 = n_2 - 1 = 5$

Using SPSS from Table 4:

$$s_1^2 = 12.754$$

$$s_2^2 = 7.535$$

where:

s_1^2 = Variance of human capital approach

s_2^2 = Variance of Log – Logistic values

$$F = s_1^2/s_2^2 = 1.693$$

From the F distribution table and calculator:

$$F_{0.292}(29,5) = 1.70$$

$$F_{0.10}(29,5) = 3.18$$

$$F_{0.05}(29,5) = 4.50$$

$$F_{0.025}(29,5) = 6.23$$

$$F_{0.01}(29,5) = 9.39$$

$$F_{0.001}(29,5) = 24.83$$

$$F_{0.00}(29,5) = 3,040,870$$

Calculated F is less than critical value of F at all significance levels. Therefore, the Log – Logistic regression model is statistically same as the human capital approach of costing road traffic accident even at an extreme risk of 29.2%.

4. Discussion

In finding an appropriate distribution model for a random variable, it is most appropriate to combine all the results of Kolmogorov Smirnov test, Anderson Darling test and Chi – Squared test by finding ranking score for each of the competing distribution models. The best distribution model, as seen in this study, is the one with the lowest ranking score. It is wrong to consider the result of only one test as this will result in misleading information.

Three challenges were made in developing the mathematical model of this research. The first challenge was Log – Logistic regression analysis of response variable from 0 to more than 1, the second challenge was the inability of the model to predict the road traffic accident costs of some developing countries, the third challenge was poor linearity between the logit and the explanatory variables. The first challenge was solved by using the knowledge of mathematics to change the signs of Equations 4 and 5, and to check that the new Equations (2 and 3) are correct. The knowledge of graphs was used to solve the second challenge. The third challenge was addressed by converting the probabilistic Gross National Income loss to an empirical data using the knowledge of iterations.

The 6 existing methods of costing road traffic accident are conceptually and quantitatively different from one another and therefore it is correct to statistically find out that the model of this study is same as the existing methods only at no risk. Being that the model is statistically the same as the human capital approach means that the approach is also suitable for developing countries.

Log of odds is directly proportional to the Gross National Income loss as seen in Figure 8. It is therefore wrong to say that as cost of casualty increases log odds decreases. However, it is true to say that as amount of road safety increases, the log odds decrease. For that reason, the modified model (Equation 11) of this study has amount of road safety instead of cost of casualty. Its complementary model and prediction are shown in Figure 8, Equation 14, and Table 5.

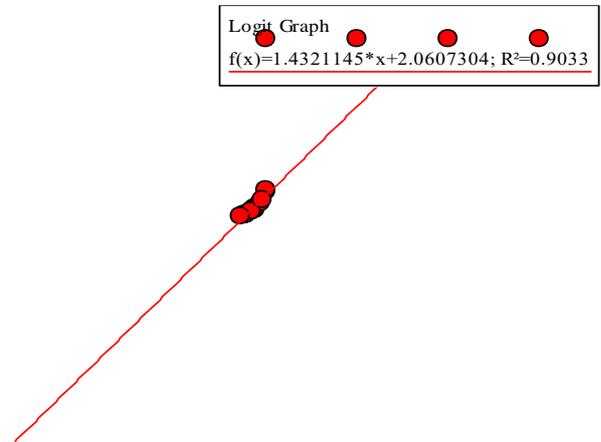


Figure 8: Log - Logistic graph

$$y = 1.4321145Y + 2.0607304 \tag{14}$$

Table 10: Complemented model prediction

| No. | Country | Y [Logit (y)] | y [Logit (Y)] |
|-----|--------------|---------------|---------------|
| 1 | Albania | -0.99 | 0.59 |
| 2 | Bangladesh | -1.13 | 0.48 |
| 3 | Brazil | 3.66 | 7.29 |
| 4 | Cameroon | -1.21 | 0.42 |
| 5 | Egypt | 5.34 | 9.70 |
| 6 | Egypt | 4.67 | 8.74 |
| 7 | Egypt | 4.60 | 8.64 |
| 8 | Egypt | 4.55 | 8.57 |
| 9 | Egypt | 2.39 | 5.49 |
| 10 | Ethiopia | -1.12 | 0.49 |
| 11 | Fiji | -1.15 | 0.46 |
| 12 | Ghana | -0.94 | 0.64 |
| 13 | India | 0.83 | 3.25 |
| 14 | Iran | 1.13 | 3.68 |
| 15 | Jordan | -0.90 | 0.69 |
| 16 | Kenya | -0.88 | 0.71 |
| 17 | Lebanon | -0.96 | 0.62 |
| 18 | Libya | -0.83 | 0.77 |
| 19 | Morocco | -0.51 | 1.52 |
| 20 | Morocco | -0.49 | 1.59 |
| 21 | Nigeria | -0.99 | 0.60 |
| 22 | Singapore | -1.15 | 0.47 |
| 23 | South Africa | 0.90 | 3.35 |
| 24 | Sudan | -0.88 | 0.70 |
| 25 | Sudan | 0.58 | 2.88 |
| 26 | Tunisia | -0.86 | 0.73 |
| 27 | Turkey | -0.71 | 0.97 |
| 28 | Uganda | -0.97 | 0.61 |
| 29 | Ukraine | -0.63 | 1.14 |
| 30 | Zimbabwe | -1.32 | 0.36 |

Comparing Figure 8 and Figure 7, and Table 5 and Table 4; it can be said that the original and modified models (Equation 9 and Equation 11 respectively) are similar. However, the modified model is far better than the original model because it was tested and found statistically the same as the human capital approach of costing road traffic accident at an extreme risk of 42.5%.

5. Conclusions and Recommendations

In this work, it can be concluded that;

- Brazil has the highest Gross National Income loss of \$142.61 billion while Tuvalu has the lowest Gross National Income loss of less than \$10 million as derived from World Bank and World Health Organization publications. The average Gross National Income loss of the developing countries is \$7.81 billion with volatility of \$20.83 billion and coefficient of variation of 267%.
- Indonesia has the highest cost of casualty of \$8.07 and highest accident severity number of 1.04 because of its high life expectancy and high casualties. Grenada, Palau, Samoa, Seychelles and St. Vincent have the lowest cost of casualty of less than \$10 million and the lowest accident severity number of less than 10,000. Solomon Island accident severity number is less than 10,000 but its cost of casualty is \$20 million. The average cost of casualty in developing countries is \$0.98 billion with volatility of \$1.71 billion and coefficient of variation of 174%. The average accident severity number of developing countries is 0.13 million with eccentricity of 0.22 million and coefficient of variation of 169%.
- In the regression analysis process, goodness of fit is an additional method of building a mathematical model to the reviewed best subsets regression and the stepwise regression methods. It showed that Log – Logistic regression analysis is suitable for this study and the developed model and its subsidiary model are;

$$Y = 0.013x_5 - 0.019x_7 + 2.280x_3 - 2.060$$

$$y = 1.4321145Y + 2.0607304$$

where:

- Y = Transformed Gross National Income loss in logit
- y = Gross National Income loss in dollars per year
- x_5 = Life expectancy in years
- x_7 = Amount of road safety
- x_3 = Accident severity number with no unit

- The accident costs predicted by the model with its subsidiary model were compared to the same accident costs obtained by the existing six methods of costing road traffic accidents. It was found out that the model and its subsidiary model are only statistically the same as all the other methods of costing road traffic accident at no risk (i.e. at 0% significance level). However, it was discovered that the models are statistically the same as the human capital approach of costing road traffic accident even at an extreme risk of 42.5% (i.e. at 42.5% significance level).
- The required mathematical model in this study has been developed using life insurance approach and Log – Logistic regression analysis.

It is recommended that;

- The model could be used by road safety stakeholders for decision making to invest in road safety of developing countries.
- The World Bank to be annually publishing list of developing countries by cost of road traffic accident calculated based on this study model with its complementary model.
- A similar study could be carried out for developed countries to enable the World Bank publish annual list of the world countries by cost of road traffic accident in terms of Gross National Income loss.
- Road traffic accident cost table of this research model and its complementary model could be developed.

- Road traffic accidents to be fully and correctly recorded and reported to the World Bank.

Acknowledgements

The authors are grateful to Prof. S. P. Ejeh, Prof. A. O. Eberemu and Dr. J. M. Kaura for their valuable questions during the research work presentations. Gratitude is extended to Prof. H. Alhassan for being external examiner of this work.

References

- Abdelwahab, H. T., & Abdel-Arty, M. A. (2001). *Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections*. Retrieved from <https://journals.sagepub.com> doi:10.3141/1746-02
- Abdelwahab, H. T., & Abdel-Arty, M. A. (2004). Predicting Injury Severity Levels in Traffic Crashes: A modeling Comparison. *Journal of Transportation Engineering*, 130 (2). Retrieved from <https://ascelibrary.org>
- Abojaradeh, M., Khaled, S., Msallam, M., & Jrew, B. (2017). Analysis and Evaluation of Traffic Accidents for Principle (sic) Urban Streets in Arbil City Iraq. *Diyala Journal of Engineering Science*, 10 (01), 118 – 131. doi: 10.24237/djes.2017.10111
- Abojaradeh, M., & Jrew, B. (2013). Development of Statistical Prediction Models to Reduce Fatal and Injury Traffic Accidents. *Journal of Environment and Earth Science*, 3 (11), 54 – 65. Retrieved from <https://www.researchgate.net>
- Abojaradeh, M. (2013). Traffic Accidents Prediction Models to Improve Traffic Safety in Greater Amman Area. *Journal of Civil and Environmental Research*, 3 (02), 87 - 101. Retrieved from www.zu.edu.jo
- Age at First Marriage. (2017). In *Wikipedia*. Retrieved from <https://en.wikipedia.org>
- Benson, G. B. (1996). *Cost Modeling for Road Accident in Nigeria: Case study of Kaduna State (Unpublished Master's Thesis)*. Ahmadu Bello University Zaria, Nigeria. Retrieved from <http://kubanni.abu.edu.ng>
- Central Intelligent Agency (2013). Inflation rates. Retrieved from <https://www.cia.gov>
- Central Intelligent Agency (2013). Literacy. Retrieved from <https://www.cia.gov>
- Centers for Disease Control and Prevention. (2019). *Road Traffic Injuries and Deaths - A Global Problem*. Retrieved from <https://www.cdc.gov>
- Developing country. (2021). In *Wikipedia*. Retrieved from <https://en.wikipedia.org>
- Fletcher, J. P., Baguley, C. J., Sexton, B., & Done, S. (2006). *Road Accident Modeling for Highway Development and Management in Developing Countries*. Department for International Development, United Kingdom.
- Garba, A. (2014). *Cost Evaluation of Road Traffic Accident in Nigeria: Case study of Abuja to Lokoja Road (Unpublished Master's Thesis)*. Ahmadu Bello University Zaria, Nigeria. Retrieved from <http://kubanni.abu.edu.ng>
- Garber, N. J., & Hoel, L. A. (2009). *Traffic and Highway Engineering* (4th ed.). Cengage Learning, United States of America.
- Gupta, S. V. (2009). *Derived Quantities and Their Units*. In Units of Measurement. Springer Series in Material Science, 122. Springer, Berlin, Heidelberg. Retrieved from https://doi.org/10.1007/978-3-642-00738-5_6
- Hejazi, R., Shamsudin, M. N., Radam, A., Abdulrahim, K., Ibrahim, Z. Z., & Yazdani, S. (2012). Estimation of Traffic Accident Costs: A Prompted Model. *International Journal of Injury Control and Safety Promotion*, 20 (2). doi: 10.1080/17457300.2012.720578
- Hills, P.J., & Jones-Lee, M.W. (1983). The role of safety in highway investment appraisal for developing countries. *Accident Analysis & Prevention*, 15, 355-369.
- International. (2014). *Road Deaths Driving to an Early Grave*. Retrieved from <https://www.economist.com>
- International Statistical Institute (2015). List of Developing Countries. Retrieved from <https://isi-web.org>
- Iqbal, S., & Zafar, A. (2017). *Determinant of Frequent Road Accidents in Megacities: A case study of Karachi*. Retrieved from <https://core.ac.uk>
- Jacobs, G., Aeron-Thomas, A., & Astrop, A. (2000). Estimating Global Road Fatalities (TRL Report 445). Crowthorne, Transport Research Laboratory.

- Julius, N. O., Anthony, K. W. and Anthony, G. W. (2015). Modeling Road Traffic Accident Injuries in Nairobi County: Model Comparison Approach. *American Journal of Theoretical Science*, 4 (3), 178. doi: 10.11648/j.ajtas.20150403.24
- Johansen, I. (2007). Graph (Version 4.3.0.384). [Computer Software]. Retrieved from <https://graph.informer.com>
- Kennedy (2002). *Department of Economics Discussion papers*. Simon Fraser University Canada.
- Odero, W., Garner, P. & Zwi, A. (1997). Road traffic injuries in developing countries: A comprehensive review of epidemiological studies. *Trop. Med. Int. Health*, 2 (5), 445–460.
- Richard, R., & Bent, T. (2004). *Road Engineering for Development (2nd Ed.)*. Sadik-Khan, J., & Sololonow, S. (2020). *The Global Traffic Death Crisis*. Retrieved from <https://www.foreignaffairs.com>
- United Nations (2013). Population. Retrieved from <https://en.wikipedia.org>
- World Bank. (2010). *Metadata Glossary*. Retrieved from <https://databank.worldbank.org>
- World Bank. (2013). *List of Developing Countries*. Retrieved from <https://richmondvale.org/>
- World Health Organization. (2011). *World Health Rankings; Live Longer Live Better*. Retrieved from www.worldlifeexpectancy.com/
- World Health Organization. (2015). *Global status report on road safety 2015*. Retrieved from <https://www.who.int>
- World Health Organization. (2018). *Road traffic injuries*. Retrieved from <https://www.who.int>
- World Health Organization. (2018). *World Health Rankings; Live Longer Live Better*. Retrieved from <https://www.worldlifeexpectancy.com>
- World Health Organization (2013). *Death Rates per 100,000 People*. Retrieved from <http://who.int>
- World Health Organization (2015). *Life Expectancy*. Retrieved from <http://www.who.int>
- World Health Organization. (2015). *World Health Rankings; Live Longer Live Better*. Retrieved from www.worldlifeexpectancy.com