Original Research   **Open Access**

Modeling COVID-19 Cases in Nigeria Using Some Selected Count Data Regression Models

Samuel Olorunfemi Adams (Corresponding Author)
Department of Statistics, University of Abuja, Abuja, Nigeria
Email: samuel.adams@uniabuja.edu.ng

Muhammad Ardo Bamanga
Department of Mathematical Sciences, Kaduna State University, Kaduna, Nigeria

Samuel Olayemi Olanrewaju
Department of Statistics, University of Abuja, Abuja, Nigeria

Haruna Umar Yahaya
Department of Statistics, University of Abuja, Abuja, Nigeria

Rafiu Olayinka Akano
Department of Statistics, University of Abuja, Abuja, Nigeria

Abstract

COVID-19 is currently threatening countries in the world. Presently in Nigeria, there are about 29,286 confirmed cases, 11,828 discharged and 654 deaths as of 6th July 2020. It is against this background that this study was targeted at modeling daily cases of COVID-19's deaths in Nigeria using count regression models like; Poisson Regression (PR), Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) model. The study aim at fitting an appropriate count Regression model to the confirmed, active and critical cases of COVID-19 in Nigeria after 130 days. The data for the study was extracted from the daily COVID-19 cases update released by the Nigeria Centre for Disease Control (NCDC) online database from February 28th, 2020 – 6th, July 2020. The extracted data were used in the simulation of Poisson, Negative Binomial, and Generalized Poisson Regression model with a program written in STATA version 14 and fitted to the data at a 5% significance level. The best model was selected based on the values of -2logL, AIC, and BIC selection test/criteria. The results obtained from the analysis revealed that the Poisson regression could not capture over-dispersion, so other forms of Poisson Regression models such as the Negative Binomial Regression and Generalized Poisson Regression were used in the estimation. Of the three count Regression models, Generalized Poisson Regression was the best model for fitting daily cumulative confirmed, active and critical COVID-19 cases in Nigeria when overdispersion is present in the predictors because it had the least -2log-Likelihood, AIC, and BIC. It was also discovered that active and critical cases have a positive and significant effect on the number of COVID-19 related deaths in Nigeria.

Keywords: Akaike information criterion (AIC); Bayesian information criterion (BIC); COVID-19; Negative binomial regression; Poisson regression.

1. Introduction

COVID-19 pandemic is the defining global health crisis of our time and the greatest challenge we have faced since World War Two. Since its emergence in China late last year, the virus has spread to every continent except Antarctica. Cases are rising daily in Africa, America, and Europe. Countries are racing to slow the spread of the virus by testing and treating patients, carrying out contact tracing, limiting travel, quarantining citizens, and canceling large gatherings such as sporting events, concerts, and schools. The pandemic is moving like a wave, one that may yet crash on those least able to cope. But COVID-19 is much more than a health crisis because it is stressing every one of the countries it touches, it has the potential to create devastating social, economic, and political crises that will leave deep scars [1].

News broke on Feb 27, 2020, that an Italian citizen was Nigeria's first case of Coronavirus Disease 2019 (COVID-19) [2]. The individual had landed at Lagos airport 2 days earlier on a flight from northern Italy and had subsequently traveled from Lagos to Ogun State, western Nigeria, where he became ill and was promptly isolated. He was treated for mild symptoms of COVID-19 in a hospital in Lagos. Upon identifying the index case, National Emergency Operations Centres were immediately activated to trace his contacts. By March 9, 2020, 27 suspected cases had been identified across five states (Edo, Lagos, Ogun, Federal Capital Territory, and Kano), of which two were confirmed to be positive (ie, the index case and a contact), with no deaths. 1,216 contacts were linked to the index case, 136 of whom are being followed up. Similar to COVID-19, the Ebola epidemic of 2014 was imported through Lagos airport. Within weeks, 19 people were diagnosed with Ebola across two states of Nigeria, Lagos and Rivers State of whom eight died of Ebola. The overstrained infrastructure in the dense population of Lagos, and the fact that it is a major regional transit hub for air, land, and sea transport created the perfect conditions for the spread

of Ebola. Nevertheless, Nigeria's aggressive and coordinated response successfully controlled the Ebola epidemic [3]. In modeling the risk of COVID-19 importation from China it was discovered that the ability of African countries to manage the local transmission of COVID-19 after importation hinges on implementing stringent measures of detection, prevention, and control [4]. Nigeria demonstrated its ability through intensifying its preparedness against COVID-19 importation, drawing on recent successes in controlling polio and Ebola epidemics [5, 6]. These experiences strengthened the health system's capacity to rapidly deploy high-quality surveillance and temperature screening at airports using equipment acquired during the Ebola epidemic; collect passengers' contact details and interview those arriving from COVID-19 hotspots, and issue travel bans [7].

However, unlike polio and Ebola, for which vaccines now exist, COVID-19 has neither a vaccine nor an approved treatment [8]. Moreover, Nigeria was rated as vulnerable to exposing huge populations to COVID-19 (potentially 200 million citizens), with a moderate capacity to control the outbreak [4]. This assessment questions Nigeria's capacity to provide sufficient bed space and associated clinical care to support those who could need isolation and quarantine if local cycles of transmission of COVID-19 occur in the country.

The research on analyzing, modeling, and forecasting the novel COVID-19 pandemic domestically and internationally has become a hot topic currently. Some of the current researches conducted include; Modelling and Forecast the number of cases of the COVID-19 pandemic with the curve estimation models like; Box-Jenkins (ARIMA) and Brown/Holt linear exponential smoothing to the number of COVID 19 epidemic cases in selected countries of G8 countries, Germany, United Kingdom, France, Italy, Russian, Canada, Japan, and Turkey, [9]. A new proposed model was used to predicts the COVID-19 epidemic and implementation in Italy [10]. Mathematical modeling was applied to the dynamics of a novel coronavirus (2019-nCoV) in Wuhan-China [11]. Mathematical modeling was applied to COVID-19 transmission mitigation strategies in Ontario Canada [12]. Some curve estimation statistical models and estimators were applied to the main factors affecting the spread of COVID-19 in Nigeria [13]. Time series was applied to model COVID-19 transmission in Mainland China as its relationship to temperature and humidity [14]. Code was presented for predicting COVID-19 cases by least-squares and fitting of the Logistic models [15]. Some parameters were used to calibrate the parameters of the SIRD model on the reported COVID-19 cases in Hubei region, China, the selected model was used to forecast the evolution of the outbreak at the epicenter for three weeks ahead [16]. A comprehensive comparison was carried out on COVID-19 cases using some mathematical model between Turkey and South Africa [17]. It is against this background that this study attempts to model the daily cumulative active, critical and confirmed COVID-19 cases as it influences the number of reported deaths in Nigeria between 28th of February to 6th of July 2020, using count regression models like; Poisson Regression (PR), Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) models. This paper is organized in the following way. This paper is organized into four sections. In section 2, the methodology used for modeling is outlined. In section 3, the results and discussions for the study are presented while section 4 presents the conclusions.

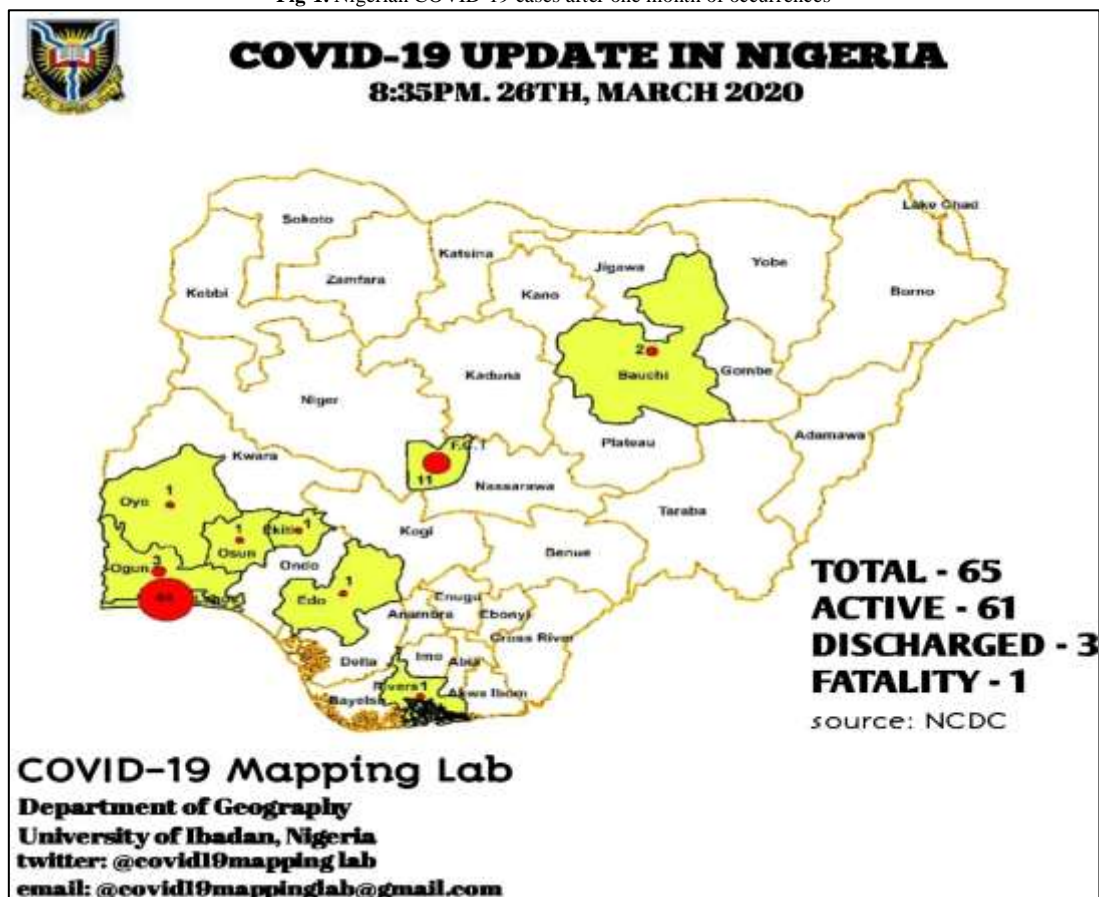**Fig-1.** Nigerian COVID-19 cases after one month of occurrences

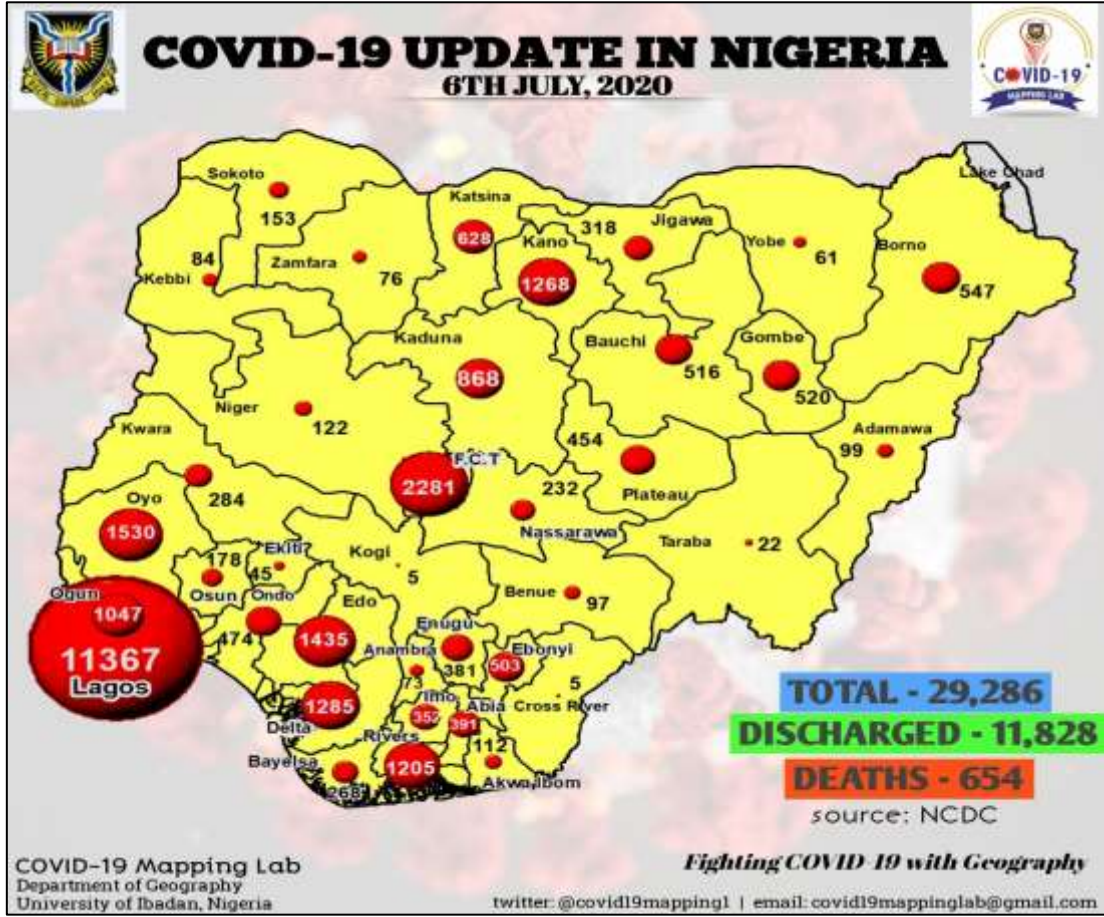**Fig-2.** Nigerian COVID -19 cases after four months of occurrences

Fig-3. Plots of cumulative (a) Confirmed cases, (b) Active cases, (c) Number of deaths, (d) Critical cases after 130 days
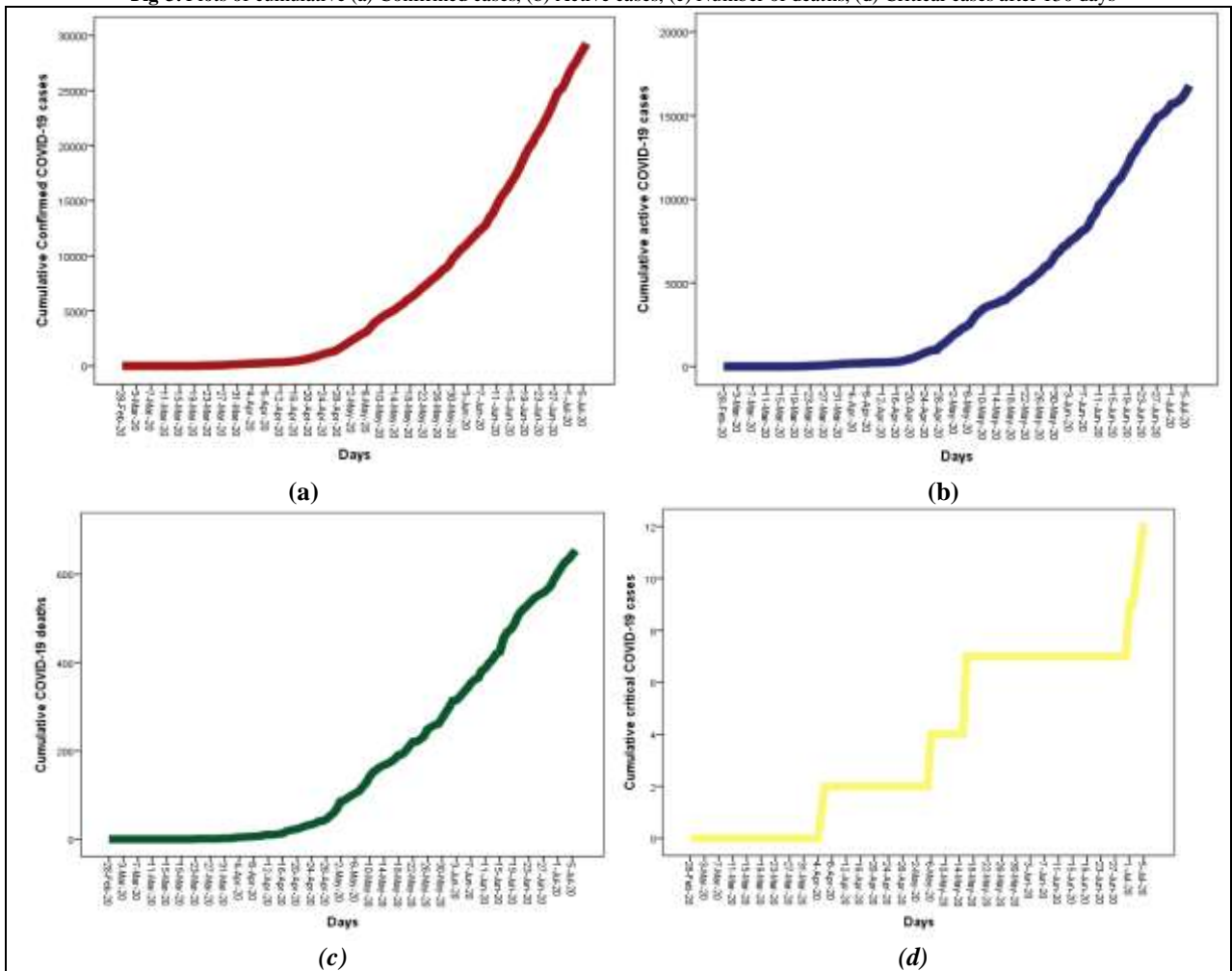

(a)

(b)

(c)

(d)

Fig. 1 presents the total number of confirmed COVID-19 cases after one month while Fig. 2 presents the total confirmed case in Nigeria after one hundred and thirty (130) days. It is evident from the two figures that COVID-19 cases had spread and increased exponentially across the states in Nigeria. At the beginning of the pandemic in Nigeria, only Lagos and Abuja had COVID-19 cases, but as of today, no state in Nigeria is free of COVID-19 while Fig. 3 presented the plot of the cumulative COVID-19 confirmed, active, death, and critical cases after one hundred and thirty (130) days of the pandemic in Nigeria. The seriousness of this pandemic becomes evident from Fig. 3. It was observed that from the 3rd of June 2020, the daily new cases of COVID-19 increased to three hundred (300) and above cases, and on the 27th of June, Nigeria record its highest ever confirmed COVID-19 cases of seven hundred and seventy-nine (779). As the daily cases of COVID-19 continue to increase, the daily number of death also continue to increase and the highest recorded case of COVID-19 was on the 16th of June, where Nigeria recorded 31 death as a result of COVID-19. Meanwhile, the World Health Organization has listed Nigeria to be among thirteen (13) other African countries with high-risk for the spread of the virus [18]. Apart from the first measure of the Federal Government of Nigeria (FGN) to strengthen surveillance at Enugu, Lagos, Rivers, Kano, and FCT International Airports on the 28th January 2020. The Nigeria Government on 31st January 2020, set up a group known as Coronavirus Preparedness Group to militate against its should incase it spread to Nigeria [6]. Other measures taken by FGN and various State Governments include the establishment of Presidential Task Force (PTF), suspension of all activities and religious gatherings, indefinite closure of public and private schools/institutions, an extension of the travel ban to some countries, suspension of the operation of Nigerian Railway Corporation, closing of borders, shops, markets, motor parks, offices, restriction of intra-states and inter-states movements and traveling out of the country. However, few states have recently relaxed the lockdown due to difficulties encountered by their citizens [6].

2. Methodology
2.1. Source of Data
The data for the study was extracted from the daily COVID-19 cases update released by the Nigeria Centre for Disease Control (NCDC) online database from February 28^{th,} 2020 – July 6^{th,} 2020, making it a total of 130 data points for total death, confirmed, active and critical COVID-19 cases [2]. The data extracted were used in the simulation of the selected count regression models with a programme written in STATA version 14.

2.2. Model Specification
In line with the specific objective of this study, the models adopted for this study are Poisson Regression, Generalized Poisson Regression and Negative Binomial Regression. The count data are the daily cumulative deaths which are considered as the dependent variable whose outcome from cumulative active, critical, and confirmed cases of COVID-19 are the independent variables. All the variables considered are continuous count data.

2.2.1. Poisson Regression (PR)
Poisson regression is a nonlinear regression analysis of the Poisson distribution used to predict a dependent variable that consists of discrete or count data given one or more independent variables. The predicted variable is called the dependent variable (or sometimes the response, outcome, target, or criterion variable). The variables used to predict the value of the dependent variable are called the independent variables (or sometimes the predictor, explanatory, or regressor variables) [19, 20]. Poisson regression is said to contain overdispersion if the variance is greater than the value of the mean value. Overdispersion has the same impact with the assumption that if the offense discrete data occurred over dispersion but still used Poisson regression, the parameter estimates of the regression coefficients remain consistent but not efficient. It is used basically when facing a problem whereby the outcome of the random process can take count values only. One of the distributions that satisfy such criterion comes from the family of exponential distribution which is the Poisson distribution [21]. Let Y be a random variable (the rate at which COVID-19 death occurs) and let be the outcomes of the cases be an event. The variable Y is said to follow a Poisson distribution with parameter $\lambda > 0$ if the probability function is given by;

$$p(Y = y) = \frac{e^{-\lambda}\lambda^n}{y!} \qquad (1)$$

Where $n = 1,2,3$ is the number of occurrences of an event and λ is defined as $\lambda = E[Y]$. One of the useful properties of the Poisson distribution is that the variance depends on the mean and also the variance is equal to the mean. The Generalized Linear Model (GLM) can be stated as thus:

$$g_i = \beta_0 + \beta_1 X_{i1+\ldots} + \beta_k X_k \qquad (2)$$

The first function describes how the mean, $E[Y_i] = \lambda_i$ which depends on the linear predictor $\varphi(\lambda_i) = g_i$ while the second functions describe how the variance, $Var(Y_i)$ depends on the mean

$Var(Y_i) = \sigma var(\lambda_i)$. Where the dispersion of parameter σ is a constant, supposing Y_i is a Poisson distribution then; $Y_i \sim Poisson (\lambda_i)$, $E[Y_i] = \lambda_i$ and $Var(Y_i) = \lambda_i$, the variance function is given as; $Var(\lambda_i) = \lambda$ and the function must map from $(0, \infty)$. A natural log of the function is given as $\ell(\lambda_i) = \log_e(\lambda_i)$.

The Generalized Linear Regression Model (GLM) suggested that the linear model should be related to the response variable through linked function, the link function here is between the linear model in a design matrix and the Poisson distribution function [22]. Supposing a linear regression model given as thus:

$$Y_i = \beta_i X_i \, x\varepsilon i \qquad (3)$$

Where X is an $nx(k + 1)$ vector of independent variables of predictors, and a column of $I's\beta$ is a $(k + 1)$ by 1 vector of unknown parameters and ε is an $n \: x \: 1$ vector of random error terms with mean zero. Therefore;

$$E(Y/X) = X_\beta \tag{4}$$

Recall that the Generalized Linear Models, where the link function and its transport Y as;

$$G(y) = log_e(y) \tag{5}$$

Therefore, this can be written in more concise form as;

$$Log_e \: E(Y/X) = X\beta \tag{6}$$

Thus, given a Poisson regression model with parameter β and its input vector X, the predicted mean of the associated Poisson distribution is given as;

$$E(Y/X) = e^{XB} \tag{7}$$

Suppose Y_i are independent observation with corresponding values X_i as the predictor variable, then the parameter β can be estimated using the Maximum Likelihood Method. The model expressed in equation (7) can be estimated by numerical methods and this is done using the logarithmic transformation of the conditional expectation of the dependent and independent variables [22]. Furthermore, the probability surface of the Maximum Likelihood Estimation of Poisson regression models is always convex form such that Newton-Raphson of the gradients–based methods are used as an appropriate estimation technique. Therefore, suppose Y_i is a random variable and it takes non-negative values such that $i = 1, 2, \ldots, n$, where n is the number of observations. Since y_i follows a Poisson distribution, the Probability Mass Function (PMF) is given as; $P(Y = y_i) = \frac{\lambda_i^n e^{-\lambda}}{y_i!}$ where ; $y_i = 0, 1, 2\ldots$ With mean and variance given as; $E[y_i] = var(y_i) = \lambda$

Where the conditional mean (predicted mean) of the Poisson distribution as given in equation (4) above specified as; $E(Y/X) = e^{x\beta} = \lambda = E(y_i)$

Where it is the value of the explanatory variable $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ are unknown K – dimensional vector of regression parameters and the mean of the predicted Poisson distribution is given as $E(Y/X)$ and its corresponding variance of Y_i as $var(Y/X)$.

The Poisson Regression which was fitted to the number of deaths in Nigeria, λ was expressed for the three (3) predictors as shown below;

$X_1 = Active \: cases$
$X_2 = Critical \: cases$
$X_3 = Confirmed \: cases$

$$Log(\theta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{8}$$

The parameter β represents the expected change in the logarithm of the mean per unit change in the predictor X_i which can be estimated by the Maximum Likelihood Estimator (MLE) method.

2.2.2. Generalized Poisson Regression (GPR)

Generalized Poisson Regression (GPR) was developed to handle the equidispersion violations assumptions on the Poisson regression model. It is mainly used to fit the overdispersed model i.e. in situations where $Var(y_i) > E(y_i)$ as well as under-dispersion, $Var(y_i) < E(y_i)$ [22]. GPR model with parameter (μ, θ) is similar to the Poisson regression models, but assumed that the components are randomly distributed to general Poisson. In the analysis of GPR, if θ is equal to 0 then the model will be the model Poisson. If θ is more than 0 then GPR models represent data containing count overdispersion case and if θ is less than 0 represent data containing under dispersion count. It was suggested that when y_i is a count response variable and it follows a Generalized Poisson distribution, the probability density function given that $i = 1, 2 \ldots. n$, then;

$$f(y_i, \mu_i^\alpha) = \left[\frac{\mu_i}{1 + \alpha\mu_i}\right]^{yi\frac{(1+ay_i)^{y_i-1}}{y_i!}} exp\left[\frac{V_i(1 + ay_i)}{1 + \alpha\mu_i}\right], y_i = 0, 1, 2, \ldots, \tag{9}$$

Where, mean is given as, $\mu_i = E(yi)$ and variances $var(y_i) = (\mu + \alpha\mu_i)^2 and \mu_i$ is refer to as the dispersion parameter [23] and [24]. Generalized Poisson distribution is a natural extension of the Poisson distribution [22]. When $\alpha = 0$, the model in equation (2) reduces to the Poisson (3), where; $Var(y_i) > E(y_i)$. When $\alpha > 0$, it means the variance $Var(y_i)$ of the distribution represents count data with over-dispersion if $\alpha > 0$, it means the variance is less than the expectation, $Var(y_i) < E(y_i)$, which simply means that the distribution represents count data with under-dispersion. The mean of the dependent variable is related to the independent variables through the link function; $\theta_i = (x, \beta)$, which is a simple linear model. This model has the disadvantage of θ_i assuming any real value but a Poisson mean assume only count nonnegative values. To correct this problem the logarithm of the linear model is used; this gives a link log function; $Log(\theta_i) = (x, \beta)$. Taking the exponential of the model we have; $(\theta_i) = \exp(x, \beta)$. In the link function, $x_i \: is \: (k - 1)$ dimensional vector of explanatory variables and β is a k dimensional vector of regression parameters and α is a dispersion parameter. The mean and variance of y_i are given by; $E(Y_i | X_i) = \theta_i$ and $Var(Y_i | X_i) = \theta_i(1 + \alpha\theta)^2$ respectively [25] and [26]. In the Generalized Poisson Regression model, the parameters (β, α) can be estimated by taking the derivatives of the log-likelihood function of the model; this means partial differentiation with respect to β and α of the logarithm function of equation (10) below;

$$LogL(\alpha, \beta, y_i) = \sum_{i=1} \left\{ y_i log\left(\frac{\theta_i}{1 + \alpha\theta_i}\right) + (y_i - 1)log(1 + \alpha y_i) - \frac{\theta_i(1 + \alpha y_i)}{1 + \alpha\theta_i} - log(y_i!) \right\} \tag{10}$$

2.2.3. Negative Binomial Regression (NBR)

Negative Binomial distribution is used to deal with the problem of over-dispersion in count data. Over-dispersion occurs when there is the presence of statistical variability in a data set. A situation in which the theoretical population means of a model is approximately the same as the sample mean. It can be further explained that this occurs when the observed variance is higher than the variance of a theoretical model, then over-dispersion is said to occur. On the other hand, under-dispersion simply means that there was less variation in the data than predicted [27]. Over-dispersion is a very common characteristic in applied data analysis because, in practice, populations are frequently heterogeneous (non-uniform) as opposed to the assumptions implicit within widely used simple parametric models. The Negative Binomial regression model used in this study specified as thus:

$$p_r[Y_i = y] = \sum p_r \left(Y_i = y_i/\theta_i\right) f(\theta_i) d_{\theta i} = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma(y_i + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left[\frac{1}{1 + \alpha\mu_i}\right]^{\frac{1}{\alpha}} \left[\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right]^{y_i}, for\ i = 0, 1, 2, \ldots \qquad (11)$$

Where, mean is given as, $\mu_i = E(yi) = V_i\left[e^{\sum \alpha xij\beta j}\right], for\ i = 1, 2, 3, \ldots, n$ while the variance of y_i is given as; $var(y_i) = (\mu + \alpha, \mu_i)^2$

Where $\alpha > 0$ the model would be referred to as a dispersion parameter, the Poisson regression model can be regarded as a limiting model of the Negative Binomial Regression model as α approaches 0.

The Maximum Log-Likelihood function of the distribution is expressed as;

$$Log(\beta, \alpha) = \sum_i \left(\sum_{r=1}^{y_i} log(1 + ar)\right) - y_i log(\alpha) - log(y_i!) + y_i log(\alpha\mu_i) - (y_i + \alpha^{-1}) log(1 + \alpha\mu_1) \qquad (12)$$

The parameter (β, α) can be estimated by a partial differential of the Maximum Likelihood function with respect to β and α. The negative Binomial Regression does not assume the equality of the mean and variance but it corrects for overdispersion that arises when the variance is greater than the conditional mean.

2.3. Model Selection Criteria

The Akaike Information Criterion (AIC), Bayesian Information criterion (BIC), and Chi-Square -2log likelihood model selection criteria shall be applied to the three (3) count data models to determine the best model for Nigeria COVID-19 cases.

2.3.1. Akaike Information Criterion (AIC)

AIC is aimed to obtain the best-approximating model to the unknown true data generating process.

One of the most commonly used information criteria which is AIC was introduced by Akaike [28]. The idea is to select the model that minimizes the negative likelihood penalized by the number of parameters as specified by the equation

$$AIC = 2K - 2lnL \qquad (13)$$

where L is the likelihood function under the fitted model and K is the number of model parameters.

2.3.2. Bayesian Information Criterion (BIC)

BIC differs from AIC only in the first term which depends on sample size n. Models that minimize the BIC are selected. From a Bayesian perspective, BIC is designed to find the most probable model given the data, it is one of the most widely used information criteria [29]. Unlike AIC, the BIC is an estimate of the Bayes factor for two competing models. The BIC is defined by

$$BIC = ln(n) - 2lnL \qquad (14)$$

Where n refers to the random sample size.

2.3.3. Chi-square -2log Likelihood Statistic

The maximized likelihood, L, for a given model is the value of the likelihood function when the parameters are substituted with their maximum likelihood estimates, and the statistic -2logL was used to compare models. It is useful for comparing models fitted to the same set of data as the value of L depends on the number of observations in the data and -2logL is a measure of agreement between the model and the data. The larger the maximum likelihood the better the agreement between the model and observed data, but the smaller the value of -2logL the better the model.

3. Results and Discussions

3.1. Summary of COVID-19 Cases in Nigeria

The cumulative COVID-19 deaths, active, critical, and confirmed cases were fitted with a family of count models such as Poisson Regression, Negative Binomial Regression, and Generalized Negative Binomial Regression).The data spanned from the 28th of February 2020 to 6th of July 2020. The summary descriptive statistic as shown in table 1 revealed that the observed cases were 130 while the mean and standard deviation of daily death as a result of COVD-19 was ($Mean = 178.4\ and\ Std.dev. = 206.9$). Similarly, the mean and standard deviation of active cases of COVID-19 was($Mean = 4419\ and\ Std.dev. = 5234.4$). The mean and standard deviation of the critical cases of COVID-19 as presented in the result showed that ($Mean = 3.8\ and\ Std.dev. = 3.3$).

However, the mean and standard deviation of the confirmed cases of COVID-19 was ($Mean = 6847.3$ and $std. dev. = 8606.1$). From the results obtained so far, it is safe to conclude that the high mean rated indicator of the active and confirmed cases of COVID-19 in Nigeria within the period under investigation (130 days) has led to a correspondingly high number of COVID-19 related deaths.

3.2. The Omnibus test of COVID-19 cases in Nigeria

The Omnibus Test table fits is presented in table 2. It is a likelihood ratio test of whether all the independent variables (i.e. active, critical, and confirmed cases of COVID-19) collectively improve the model over the intercept-only model (i.e., with no independent variables added). The result from the table indicates that all the independent variables have a p-value of .000 (i.e., p < .05), in modeling with Poisson Regression (PR), Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR), indicating a statistically significant overall model.

3.3. Multi-Collinearity Test of COVID-19 Cases in Nigeria

The Multicollinearity test is needed to be done as an initial assumption for parameter estimation. The criteria that can be used to detect cases of multicollinearity is by computing the value of VIF. Multicollinearity occurs if the VIF value is greater than 10. The VIF value of each predictor variable is presented in Table 3. The result shows that all predictor variables have VIF values <10. Thus all the variables can be included in the subsequent analysis.

Table-1. Summary of Descriptive Statistics of COVID-19 cases in Nigeria

Variables	Observations	Total	Mean	Std. Dev.	Min.	Max.
Deaths	130	654	178.4	206.9	0	31
Active cases	130	16804	4419	5234.4	0	13865
Critical cases	130	12	3.8	3.3	1	12
Confirmed cases	130	29286	6847.3	8606.1	1	29286

Source: Researcher's Computation, 2020 using STATA Version 14

Table-2. Omnibus test of COVID-19 cases in Nigeria

Models	LR Chi-Square	d.f	P-value
Poisson Regression	16954.55	3	0.000
Negative Binomial Regression	99.865	3	0.000
Generalized Poisson Regression	272.459	3	0.000

Source: Researcher's Computation, 2020 using STATA Version 14

Table-3. VIF value of 3 predictor variables

Predictor variables	VIF	1/VIF
Active cases (x_1)	412.95	0.002422
Critical cases(x_2)	16.85	0.059347
Confirmed cases(x_3)	351.33	0.002422
Mean	260.38	

Source: Researcher's Computation, 2020 using STATA Version 14

3.4. Modelling the Number of COVID-19 Cases in Nigeria Using Poisson Regression (PR)

The confirmed death cases as a result of COVID-19 are a data count, and this type of data follows the Poisson distribution. The modeling with Poisson regression analysis is conducted to determine the factors that influence the number of COVID-19 related deaths. Maximum Likelihood Estimator (MLE) method is used to obtain the estimation of the Poisson regression model parameters as shown in Table 4 and the resulting log-likelihood, AIC and BIC values of -2140.7646, 4289.529 and 4300.999 respectively. The result presented in Table 4 also shows that the p-value of all the parameters is smaller than 0.05, implying that the parameters β_1, β_2, and β_3 have a significant effect on the model. The Poisson regression models generated by.

$$y_i = exp(3.168602 + 0.000616X_1 + 0.110749X_2 - 0.00027X_3) \tag{15}$$

Predictor variables that influence the number of COVID-19 deaths in Nigeria are the COVID-19 active cases (X_1), critical cases (X_2) and confirmed cases (X_3).

3.5. Overdispersion Test

In the Poisson regression modeling, there is an assumption about equidispersion that the mean value and the variance must be equal. However, this assumption is rarely met because overdispersion is present in most cases. To detect overdispersion, the value of deviance when divided by the degree of freedom must be greater than 1 but when the division of deviance value by degree of freedom (d.f.) is less than 1 then there exist under dispersion. Table 5 shows that the value of deviance/df and Pearson chi-square/df is greater than 1 so it can be concluded on the Poisson Regression models that the number of COVID-19 cases in Nigeria occurred overdispersion. Overdispersion will lead to models with biased parameter estimates and will also have consequences on the value of the standard error estimator for the smaller (underestimate) which subsequently can lead to errors in inference for the parameters. To overcome this problem, the modeling is done using the Negative Binomial Regression (NBR) and Generalized Negative Binomial Regression (GNBR) because both methods can accommodate the dispersion parameter.

3.6. Modelling the Number of COVID-19 Cases in Nigeria Using Negative Binomial Regression (NBR)

To overcome the problem of overdispersion association with Poisson Regression modeling of count data, Negative Binomial Regression and the Generalized Poisson Regression was considered, since both models can accommodate and capture dispersion parameter in modeling counts data. The result of estimation using the Negative Binomial Regression model in table 6 shows that all the parameters are significant at the 5% level and it can be seen from the result that the p-values of the parameters are smaller than 0.05. The Negative Binomial Regression model also revealed that two coefficients of the COVID-19 cases were positive except the confirmed cases. This confirmed that in estimating the coefficient using Negative Binomial Regression, a 1% increase in active and critical cases of COVID-19 may lead to 0.15% and 0.31% increase in the number of COVID-19 related deaths in Nigeria. Also, their z-statistics (5.01 and 4.08) are greater than 2 and by the rule of the thumb, it confirmed that active and critical COVID-19 cases have a positive and significant effect on the number of deaths associated with COVID-19 in Nigeria. However, the confirmed cases have a negative coefficient (-0.00077) at the 5% level of significance which simply means that a 1% increase in confirmed cases of COVID-19 in Nigeria during the period under investigation may lead to (-0.077%) decrease in the number of death through COVID-19 in Nigeria. The z-statistic coefficient of (-4.9) shows that confirmed cases are less than 2 and by the rule of the thumb, the result indicated that confirmed cases have a negative effect of the number of COVID-19 death.

3.7. Modelling the Number of COVID-19 Cases in Nigeria Using Generalized Poisson Regression (GPR)

The result of estimation using the Negative Poisson Regression model in table 7 shows that all the parameters are significant at 5% significance level and it can be seen that the p-values of all parameters were less than 0.05. The result also revealed that the three coefficients of the COVID-19 cases were positive except the confirmed cases. This confirmed that in estimating the predictors using Generalized Poisson Regression, a 1% increase in active and critical cases of COVID-19 may lead to 0.1% and 26% increase in the number of COVID-19 related deaths in Nigeria. Also, their z-statistics values of (4.95 and 3.69) are greater than 2 and by the rule of the thumb, this confirming that active and critical COVID-19 cases have a positive significant effect on the number of deaths associated with COVID-19 in Nigeria. However, confirmed cases have a negative coefficient (-0.0003882) at the 5% level of significance which simply means that a 1% increase in confirmed cases of COVID-19 in Nigeria during the period under investigation may lead to (-0.039%) decrease in the number of death through COVID-19 in Nigeria. The z-statistic coefficient of (-3.76) shows that the confirmed cases are less than 2 and by the rule of the thumb, it implies that confirmed cases have no negative effect of the number of COVID-19 death.

Table 8 shows the models' goodness of fit criteria result for the three count data, namely; Poisson Regression (PR), Negative Binomial Regression (NBR), and Generalized Poisson Regression. The result indicated that the Generalized Poisson Regression (GPR) has the least AIC (1053.422), BIC (1116.533), and -2log likelihood (-384.1195) values and is the preferred count data model for modeling active, critical and confirmed COVID-19 pandemic cases in Nigeria.

Table-4. Poisson Regression (PR) Parameter Estimate and Performance Criteria of COVID-19 Cases in Nigeria

Parameters	Parameter estimate				Performance criteria		
	Coefficient	Std.error	Z	$p > \lvert z \rvert$	Log-likelihood	AIC	BIC
β_0	3.168602	0.02082	152.18	0.000	-2140.7646	4289.529	4300.999
β_1	0.000616	0.00002	32.46	0.000			
β_2	0.110749	0.00389	28.51	0.000			
β_3	-0.00027	0.02982	-25.80	0.000			

Source: Researcher's Computation, 2020 using STATA Version 14

Table-5. Value of Deviance and Pearson Poisson Regression Model

Criterion	Value	d.f.	Value/df
Deviance	88072.024	128	688.063
Pearson Chi-Square	69541.152	128	543.290

Source: Researcher's Computation, 2020 using STATA Version 14

Table-6. Negative Binomial Regression (NBR) parameter Estimate and Performance criteria of COVID-19 cases in Nigeria

Parameters	Parameter estimate				Performance criteria		
	Coefficient	Std.error	Z	$p > \lvert z \rvert$	Log-likelihood	AIC	BIC
β_0	1.599214	0.16575	9.65	0.000	-643.36058	1296.721	1311.059
β_1	0.001466	0.00029	5.01	0.000			
β_2	0.305694	0.07497	4.08	0.000			
β_3	-0.00077	0.00016	-4.86	0.000			

Source: Researcher's Computation, 2020 using STATA Version 14

Table-7. Generalized Poisson Regression (GPR) Parameter Estimate and performance criteria of COVID-19 cases in Nigeria

Parameters	Parameter estimate				Performance criteria		
	Coefficient	Std.error	Z	$p > \lvert z \rvert$	Log-likelihood	AIC	BIC
β_0	1.327642	0.14353	8.37	0.000	-384.1195	1053.422	1116.533
β_1	0.0010232	0.00015	4.95	0.000			
β_2	0.2603251	0.04745	3.69	0.000			
β_3	-0.0003882	0.00012	-3.76	0.000			

Source: Researcher's Computation, 2020 using STATA Version 14

Table-8. Preferred Count Data Model for COVID-19 Cases in Nigeria

Models	AIC	BIC	-2Log-likelihood
PR	4289.529	4300.99	-2140.7646
NBR	1296.721	1311.059	-643.36058
GPR*	1053.422*	1116.533*	-384.1195*

* Refers to the best-fitted model i.e. Generalized Poisson Regression

4. Conclusion

We fitted three selected count Regression models on active, critical, and confirmed cases as factors that influence the novel coronavirus disease 2019, (COVID-19) deaths in Nigeria whose first case was reported in December 2019 in Wuhan China. The data for the study was sourced from the daily cumulative COVID-19 cases update released by the Nigeria Centre for Disease Control (NCDC) online database from the February 28[th,] 2020 – July 6[th,] 2020. We presented a summary of the 130 observed cases of the cumulative death, active, critical, and confirmed COVID-19 cases in Nigeria during the period under investigation. The results obtained from the analysis revealed that the Poisson Regression could not capture over-dispersion, so other forms of Poisson Regression models such as the Negative Binomial Regression and Generalized Poisson Regression were used in the estimation. Generalized Poisson Regression (GPR) was the best model selected. This was inferred from the results generated from the model selection criteria applied which include; -2logL, AIC, and BIC. All these criteria established the Generalized Poisson Regression (GPR) as the best model because it had the smallest value of all three selection criteria.

A novel finding of this study is the positive significant influence between active and critical cases on the number of COVID-19 deaths in Nigeria. Active and critical cases of COVID-19 was found to have a positive effect on the number of COVID-19 related deaths, for NBR model: $(\beta_1 = 0.001466, p < 0.000)$ implies that 1% increase in active cases led to a 0.15% increase in the number of deaths while (0.305694, p<0.000). 1% increase in critical cases led to a 0.31% increase in the number of deaths in Nigeria. For GPR model; $(\beta_1 = 0.0010252, p < 0.000)$, implies that, 1% increase in active cases led to 0.1% increase in number of deaths while $(\beta_2 = 0.2603251, p < 0.000)$ shows that a 1% increase in critical cases led to a 26% increase in the number of deaths in Nigeria. It was also demonstrated that Generalized Poisson Regression (GPR) model is the best model to determine the factors that influence COVID-19 deaths in Nigeria when there is an indication of the presence of overdispersion in the count data.

Based on the results and findings from this study, the following recommendations were made; Generalized Poisson Regression (GPR) should be used to model the daily critical, active, and confirmed cases as factors that affect COVID-19 related death and Nigerian Government through the Presidential Task Force (PTF) on COVID-19 and Nigeria Centre for Disease Control (NCDC) should pay more attention to the critical and active cases because they are the major factors that affect COVID-19 related deaths in Nigeria.

Acknowledgments

Competing Interests Disclaimer

Authors have declared that no competing interests exist.

Reference

[1] UNDP, 2020. "COVID-19 Pandemic: Humanity needs leadership and Solidarity to defeat the Coronavirus."

[2] Nigeria Centre for Disease Control NCDC, 2020. "COVID-19 outbreak in Nigeria: situation report." Available: http://covid19.ncdc.gov.ng

[3] MacLean, O. A., Orton, R., Singer, J. B., and Robertson, D. L., 2020. "Response to on the origin and continuing evolution of SARS-CoV-2." Available: http://virological.org/t/response-to-on-the-origin–and-continuing-evolution-of-SARS-COV-2/418

[4] Gilbert, M., Pullano, G., and Pinotti, F., 2020. "Preparedness and vulnerability of African countries against importations coef COVID-19: A modeling study." *Lancet,* Available: https://doi.org/10.1016/S0140–6736(20)30411–6

[5] Musa, E., Nasidi, A., Shuaib, F., Nguku, P. M., and Vaz, R. G., 2016. "Nigeria's Ebola outbreak response: lessons for future epidemic preparedness." *Arch. Med.,* vol. 8, p. 6.

[6] Unah, L., 2020. "Nigeria will deal with this': high alert after coronavirus case. March 1." Available: https://www.aljazeera.com/news/2020/03/deal-highalert-coronavirus-case200301140838366

[7] WHO, 2020. "Coronavirus disease 2019 (COVID-19) situation report–100." Available: https://www.who.int/20200425-sitrep-96-19

[8] Wiki, 2020. "Coronavirus pandemics in Nigeria." Available: http://en.wikipedia.org/wiki/2020

[9] Yonar, H., Yonar, A., Tekindal, M. A., and Tekindal, M., 2020. "Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods." *EJMO,* vol. 4, pp. 160–5.

[10] Giordano, G., Blanchini, F., Bruno, R., Colaner, P., Filippo, A. D., Matteo, A. D., and Colaneri, M., 2020. "Modeling COVID-19 Epidemic and Implementation of population-wide interventions in Italy." *Nat Med,* vol. 26, pp. 855-860.

[11] Khan, M. A. and Atangana, A., 2020. "Modeling the dynamics of novel coronavirus (2019-n-Cov) with fractional derivative." *Alexandria Engineering Journal,* Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7148705/

[12] Tuite, A. R., Fisman, D. N., and Greer, L. A., 2020. "Mathematical modeling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada." *CMAJ,* vol. 192 (19), pp. E497-E505.

[13] Ayinde, K., Lukman, A. F., Rauf, I. R., Alabi, O. O., Okon, C. E., and Ayinde, O. E., 2020. *Modeling Nigerian Covid-19 cases: A comparative analysis of models and estimators*. Chaos: Solitons and Fractals.

[14] Qi, H., Shuang, Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., Su, Q., Wang, W*., et al.*, 2020. "Covid-19 transmission in mainland China is associated with temperature and humidity: A time series analysis." *Science of the Total Environment,* vol. 728, Available: https://www.sciencedirect.com/science/article/pii/S0048969720322956

[15] Ademir, X., 2020. "AC ++ code for predicting COVID-19 cases by least-squares fitting of the Logistic model." Available: https://www.researchgate.net/publication/340273721_A_C_code_for_predicting_COVID-19_cases_by_least-squares_fitting_of_the_Logistic_model

[16] Anastassopoulou, C., Russo, L., Tsakris, A., and Siettos, C., 2020. "Data-based analysis, modeling and forecasting of the COVID-19 outbreak." *PLoS ONE,* vol. 15, p. e0230405. Available: https://doi.org/10.1371/Journal.pone.0230405

[17] Atangana, A. and Araz, S. I., 2020. "Mathematical Model of COVID-19 spread in Turkey and South Africa: Theory, Methods, and Applications." Available: https://www.researchgate.net/publication/341360578 https://doi:10.1101/2020.05.08.20095588

[18] Ifijeh, M., 2020. *FG sets up coronavirus preparedness group*. This Day Newspaper.

[19] Hardin, J. W. and dan Hilbe, J. M., 2007. *Generalized linier models and extensions*. Texas: A Strata Perss Publication.

[20] McCullagh, P. and dan Nelder, J. A., 1983. *Generalized linier models*. London: Chapmann and Hall.

[21] Philip, N. and Sebastian, N., 2017. "Application of poisson regression on traffic safety. Degree project, in second level mathematical statistics stockholm." Available: https://www.diva-portal.org/smash/get/diva2:816402/FULLTEXT01.pdf

[22] Nwankwo, C. H. and Nwaigwe, G. I., 2016. "A statistical model of road traffic crashes data in Anambra State, Nigeria: a Poisson regression approach." *European Journal of Statistics and Probability,* vol. 1, pp. 41-60.

[23] Famoye, F., 1993. "Restricted generalized poisson regression model. Communications in statistics." *Theory and Methods,* vol. 22, pp. 1335-1354.

[24] Wang, W. R. and Famoye, F., 1997. "Modeling household fertility decisions with generalized poisson regression." *Journal of Population Economics,* vol. 10, pp. 273-283.

[25] Sadia, F., 2013. "Performance of generalized poisson regression model and negative binomial regression model in the case of over-dispersion " *Count Data,* pp. 558–563.

[26] Ismail, N. and Jemain, A. A., 2007. *Handling overdispersion with negative binomial and generalized poisson regression models*. Casualty Actuarial Society Forum Winter, pp. 103–158.

[27] Shaw-Pin, M., 1993. *The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions*. Center for Transportation Analysis, Energy Division Oak Ridge National Laboratory.

[28] Akaike, H., 1973. *Information theory and extension of the maximum likelihood principle, In: B.N. Petrov and F. Csaki (eds), international Symposium on Information Theory*. 2nd ed. Budapest: Akademiai Kiado. pp. 267-87.

[29] Schwarz, G., 1978. "Estimating the dimension of a model." *Annals of Statistic,* vol. 6, pp. 461-464.