MODELING NOVEL COVID-19 PANDEMIC IN NIGERIA USING COUNT DATA REGRESSION MODELS

*David Adugh Kuhe, Enobong Francis Udoumoh and Ukamaka Lawrensia Ibeajaa

Department of Statistics, Joseph Sarwuan Tarka University, Makurdi-Nigeria

*Corresponding authors' email: davidkuhe@gmail.com   Phone: +2348064842229

ABSTRACT

This study aimed to model COVID-19 daily cases in Nigeria, focusing on confirmed, active, critical, recovered, and death cases using count data regression models. Three count data regression models-Poisson regression, Negative Binomial regression, and Generalized Poisson regression were applied to predict COVID-19 related deaths based on the mentioned variables. Secondary data from the Nigeria Centre for Disease Control (NCDC) between February 29, 2020, and October 19, 2020, were used. The study found that Poisson Regression could not handle over-dispersion inherent in the data. Consequently, Negative Binomial Regression and Generalized Poisson Regression were considered, with Generalized Poisson Regression identified as the best model through performance criteria such as -2 log likelihood (-2logL), Akaike information criterion (AIC), and Bayesian information criterion (BIC). The study revealed positive and significant impacts of confirmed, active, and critical cases on COVID-19 related deaths, while recovered cases had a negative effect. Recommendations included increased attention to confirmed, active, and critical cases by relevant authorities to mitigate COVID-19-related deaths in Nigeria.

Keywords: Betacoronavirus, Count Data, Exponential Family, Over-dispersion, Nigeria

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by the novel coronavirus SARS-CoV-2. The origins of the virus are believed to be zoonotic, meaning it likely originated in animals before transmitting to humans (Cheng and Shan, 2020). The first known cases of COVID-19 were reported in December 2019 in the city of Wuhan, Hubei province, China. Many early cases were linked to a seafood market in Wuhan, suggesting a possible zoonotic origin. The market also sold live wild animals, raising concerns about the virus's transmission from animals to humans (Cheng and Shan, 2020).

Scientists initially suspected that the virus might have originated in bats and then jumped to humans, possibly through an intermediate host species (Hassan *et al*., 2020). Some early studies suggested a link to pangolins, a type of mammal, as a potential intermediary host, though this has not been definitively proven (Shereen *et al*., 2020). The SARS-CoV-2 virus is a betacoronavirus, similar to the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) that caused the SARS outbreak in 2002-2003. Genetic analysis indicates that SARS-CoV-2 shares a high degree of similarity with coronaviruses found in bats (Hassan *et al*., 2020).

The virus quickly spread globally, leading to a pandemic. The World Health Organization (WHO) declared COVID-19 a Public Health Emergency of International Concern on January 30, 2020, and later declared it a pandemic on March 11, 2020 (WHO, 2020). International efforts, including studies conducted by the World Health Organization, have been underway to investigate the origins of SARS-CoV-2. These investigations involve collaboration between scientists from various countries and aim to understand how the virus emerged and whether it involved an intermediary host species (WHO, 2020).

Nigeria confirmed its first case of COVID-19 on February 27, 2020. The index case was an Italian citizen who arrived in Lagos, Nigeria's largest city, from Milan, Italy. Following the confirmation of the first case, the Nigerian government implemented various measures to curb the spread of the virus (NCDC, 2020). These measures included travel restrictions, quarantine protocols, and public health awareness campaigns.

Subsequent cases in Nigeria were primarily linked to travel or contact with confirmed cases. However, community transmission soon became a concern, prompting the government to escalate response efforts.

To contain the spread of the virus, Nigeria implemented lockdowns and restrictions in major cities. These measures included the closure of schools, businesses, and restrictions on movement. Nigeria scaled up testing and surveillance efforts to identify and isolate cases promptly. Testing centers were established across the country to enhance diagnostic capabilities. Nigeria collaborated with international organizations such as the World Health Organization (WHO) and received support in terms of expertise, medical supplies, and equipment. As COVID-19 vaccines became available, Nigeria initiated vaccination campaigns to immunize the population against the virus. The vaccination efforts aimed to achieve widespread coverage and reduce the impact of the disease.

Modeling COVID-19 cases is crucial in public health and epidemiology for several reasons. It helps researchers and officials understand the virus's spread, considering factors such as transmission rates and intervention impacts. Modeling anticipates healthcare resource needs, aiding preparedness and allocation of resources like hospital beds, ventilators, PPE, and staff. Policymakers rely on models for informed decisions on interventions, including lockdowns, social distancing, travel restrictions, and vaccination campaigns. Models guide vaccine distribution planning by identifying priority populations and estimating doses required. They monitor the pandemic's progression, track intervention effectiveness, and provide early warnings for timely responses. Models support effective public communication, explaining health recommendations and projecting potential consequences. Additionally, modeling contributes to the development of treatments and vaccines, offering a framework to understand intervention impacts. Collaboratively, models facilitate international cooperation, enabling a coordinated global response to the pandemic. It is crucial to acknowledge that models are simplifications with inherent uncertainties but remain invaluable for decision-making in rapidly evolving health crises like COVID-19.

Count data regression models, such and Poisson regression, Negative Binomial regression and Generalized Poisson regression, can be valuable in modeling COVID-19 cases due to their ability to handle discrete count data. These models are specifically designed for variables that represent the number of occurrence of an event in a fixed unit of time, space, or other well-defined intervals such as disease cases, traffic accidents or customer arrivals. They provide flexibility in handling discrete, non-negative data with a skewed distribution, allowing for variations in the data that may not be adequately captured by traditional linear regression models (Nwankwo and Nwaigwe, 2016)

The provided text discusses various studies that utilized statistical models as well as Poisson-related models to analyze and forecast infectious disease dynamics, particularly focusing on COVID-19 cases in different regions.

Babatunde and Igboeli (2020) conducted research aiming to understand the COVID-19 virus spread dynamics by employing Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN). The objective was to identify the most suitable model for forecasting future occurrences of the pandemic, especially anticipating the second wave. The study revealed that both linear and nonlinear prediction models effectively captured the virus trend in Nigeria. ARIMA demonstrated over 97% accuracy over a 120-day period, while ANN yielded approximately 98.01% accuracy in certain states. The research concluded that ARIMA or ANN could accurately predict future waves of the virus and similar epidemics.

Ibrahim and Oladipo (2020) undertook a study focusing on analyzing COVID-19 spread in Nigeria using statistical models and data from the NCDC. Their research aimed to establish a predictive model to guide health interventions and mitigate virus spread. Daily spread data from February 27 to April 26, 2020, were utilized to build an autoregressive integrated moving average (ARIMA) model using R software. The study conducted stability analysis, stationarity tests, parameter tests, and model diagnostics. Based on AICc model selection criteria, the ARIMA (1,1,0) model was chosen, projecting a steep increase in COVID-19 spread in Nigeria over a ten-day period.

Adams *et al.* (2020) studied COVID-19 cases in Nigeria using Poisson, Negative Binomial, and Generalized Poisson Regression models. They found that the Generalized Poisson Regression was the best model, highlighting the positive impact of active and critical cases on COVID-19-related deaths. Barria-Sandoval *et al.* (2021) modeled COVID-19 cases and deaths in Chile using ARIMA, Exponential Smoothing, and Poisson models. They determined ARIMA as the best model for confirmed cases and a damped trend method for predicting deaths.

Agosto and Giudici (2020) favoured Poisson autoregressive models over exponential growth models for studying COVID-19 cases in South Korea, Iran, and Italy, citing superior predictive performance. Nasution *et al.* (2021) used Poisson Autoregression to predict COVID-19 cases in Jakarta, finding it more accurate than ARIMA, Exponential Smoothing, BATS, and Prophet models. Tawiah *et al.* (2021) investigated COVID-19 deaths in Ghana using zero-inflated models. They found the zero-inflated negative binomial autoregressive model outperformed others, predicting a sharp rise and fall in deaths. Alzahrani (2022) applied log-linear Poisson autoregressive and ARIMA models to study COVID-19 in Saudi Arabia, concluding that the former provided better forecasting and could be applied comprehensively.

Adewole *et al.* (2021) formulated a mathematical model to comprehend COVID-19 dynamics in Nigeria, incorporating key compartments and parameters. They obtained the basic reproduction number and analyzed the stability of disease-free equilibrium. Using data from the Nigeria Centre for Disease Control, they calibrated the model and estimated its key parameters. Sensitivity analysis explored parameter influence on disease control, and optimization using Pontryagin's maximum principle identified time-dependent intervention strategies to suppress virus transmission. Numerical simulations explored optimal control solutions, highlighting the necessity of stringent intervention efforts for rapid disease suppression.

Zhao *et al.* (2022) proposed an autoregressive count data model for predicting COVID-19 cases in Sweden. They recommended specific lag periods based on population density and intervention inclusion for accurate predictions. Giudici *et al.* (2022) introduced Bayesian time-dependent Poisson autoregressive models to estimate the impact of policy measures on COVID-19 cases in Italy and the United States. Their results indicated a significant reduction in disease counts due to policy interventions, including closures and vaccine distribution.

Overall, these studies highlight the versatility and effectiveness of Poisson-related models in understanding and forecasting infectious disease dynamics, particularly during the COVID-19 pandemic.

MATERIALS AND METHODS
Data Source
The data used in this work comprised daily confirmed, active, critical, recovered and death cases due to COVID-19 in Nigeria from 29th February, 2020 to 19th October, 2020. The data was obtained as secondary data from NCDC (2020) website.

Model Specification
The models adopted for the current study are Poisson Regression (PR), Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR). The count data are the daily death cases which are considered as the dependent variable whereas the daily confirmed, active, critical, and recovered cases of COVID-19 are the independent variables. All the variables considered are continuous count data.

Poisson Regression Model
Poisson regression is a type of regression analysis used when the dependent variable is a count variable, representing the number of events occurring in a fixed interval of time or space. The Poisson regression model assumes that the counts follow a Poisson distribution, which is characterized by the mean rate of events (λ) occurring in the fixed interval (Agresti, 2007).

The mathematical form of the Poisson regression model is given by:

$$\ln(\lambda_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i \quad (1)$$

where λ_i is the expected COVID-19 related deaths for the i-th observation, $\ln \lambda_i$ is the natural logarithm of the expected COVID-19 related death, and it serves as the link function, β_0 is the intercept term, x_1 = confirmed cases, x_2 = active cases, x_3 = critical cases, x_4 = recovered cases, $\beta_1, \beta_2, \beta_3$ and β_4 are the slope coefficients corresponding to the confirmed, active, critical and recovered COVID-19 cases respectively and ε_i are the error terms which account for the unexplained variations or factors not included in the model.

The model assumes that the observed counts y_i follow a Poisson distribution with mean λ_i:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \quad (2)$$

The link function $\ln(\lambda_i)$ transforms the linear combination of the independent variables into the range of real numbers, ensuring that the predicted values are positive. The Poisson regression model expresses the logarithm of the expected count as a linear combination of predictor variables, and it is commonly estimated using maximum likelihood methods.

The Negative Binomial Regression

The Negative Binomial Regression model is a type of generalized linear model (GLM) that is used for count data when the variance is greater than the mean, which is a common characteristic of count data. The Negative Binomial Regression model is particularly suitable when there is over-dispersion in the data, meaning that the variability is higher than what would be expected under a Poisson distribution (Hilbe, 2011).

The probability mass function (PMF) of the Negative Binomial distribution is given by:

$$P(Y - y) = \frac{\Gamma(y+r)}{\Gamma(r).y!} . p^r . (1-p)^y \qquad (3)$$

where Y is the random variable representing the count, y is the observed count, r is the dispersion parameter (shape parameter) of the distribution, p is the probability of success in a single trial and $\Gamma(.)$ is the gamma function (Hilbe, 2011).

The mean μ and variance σ^2 of the Negative Binomial distribution are given by:

$$\mu = \frac{r(1-p)}{p} \text{ and } \sigma^2 = \frac{r(1-p)}{p^2} \qquad (4)$$

In the context of regression modeling, the Negative Binomial Regression model extends the Poisson Regression model by introducing a dispersion parameter r. The mean μ is related to the linear predictor η through the logarithmic link function, $\ln(\mu) = \eta$ (Hilbe, 2011).

The linear predictor η is modeled as a linear combination of predictors:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon_i \qquad (5)$$

where x_1 = confirmed cases, x_2 = active cases, x_3 = critical cases, x_4 = recovered cases, β_0 = intercept, $\beta_1, \beta_2, \beta_3$ and β_4 are the slope coefficients corresponding to the confirmed, active, critical and recovered COVID-19 cases respectively and ε_i are the error terms which account for the unexplained variations or factors not included in the model.

The dispersion parameter r is estimated along with the coefficients during the model fitting process. The Negative Binomial Regression model allows for handling over-dispersed count data and provides a flexible approach for modeling relationships between predictors and count outcomes.

The Generalized Poisson Regression (GPR)

The Generalized Poisson Regression (GPR) is a model that extends the Poisson Regression by introducing an additional parameter to account for over-dispersion (Agresti, 2007). The probability mass function (PMF) of the Generalized Poisson distribution is given by:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{(1 + \alpha y) y!} \qquad (6)$$

where Y is the random variable representing the count, y is the observed count, λ is the mean parameter and α is the over-dispersion parameter.

The mean μ of the Generalized Poisson distribution is given as $\mu = \lambda$ and the variance σ^2 is given as $\sigma^2 = \lambda(1 + \alpha\lambda)$.

In the context of regression modeling, the Generalized Poisson Regression model models the logarithm of the mean λ as a linear combination of predictors using the link function. The link function is often the logarithmic function, leading to a linear predictor η, $\ln(\lambda) = \eta$ (Agresti, 2007).

The linear predictor η is modeled as:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon_i \qquad (7)$$

All the parameters are as earlier defined.

The model parameters, including the over-dispersion parameter α, are estimated during the model fitting process. The Generalized Poisson Regression allows for flexibility in handling over-dispersed count data, and it is particularly useful when the assumption of equidispersion in the Poisson model is violated. The choice of the link function and the inclusion of predictors follow similar principles to other generalized linear models.

Model Evaluation and Diagnostic Checks

The following goodness-of-fit measures are employed to evaluate model performance and to check the adequacy of the estimated models.

Deviance statistic

Deviance is defined as the log likelihood of the final model multiplied by (-2). It is mathematically expressed as:

$$\text{Deviance} = \sum_{i=1}^{n} 2 \left[y_i \log \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right] \qquad (8)$$

where $\hat{y}_i$ is the predicted value of y_i.

Pearson chi-square

This is a goodness-of-fit measure that compares fitted values of the outcome variables with the actual values. It is computed mathematically as:

$$\mathcal{X}_p^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \qquad (9)$$

where $\hat{y}_i$ is the predicted value of y_i.

Akaike information criterion (AIC)

Akaike information criterion (AIC) is a goodness-of-fit measure defined as:

$$\text{AIC} = -2 \ln L + 2k \qquad (10)$$

Bayesian information criterion (BIC)

Bayesian information criterion (BIC) is a goodness-of-fit measure defined as:

$$\text{BIC} = \frac{-2 \ln L + k \ln(n)}{n} \qquad (11)$$

where n is the total number of observations, k is the number of model parameters and L is the likelihood function of the final model defined as:

$$L = \prod_{i=0}^{n} \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} exp \left[-\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma_i^2} \right] \qquad (12)$$

$$\ln(L) = In \left[\prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \right] - \frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma_i^2} \qquad (13)$$

Thus given a set of estimated Poisson regression models for a given dataset, the preferred model is the one with the minimum values of information criteria and maximum log likelihood value.

RESULTS AND DISCUSSION
Summary Statistics of COVID-19 Cases in Nigeria
Table 1 displays the computed summary statistics for the confirmed, active, critical, recovered, and death cases of COVID-19 in Nigeria.

Table 1: Summary Statistics of COVID-19 Cases in Nigeria

Variable	Death Cases	Confirmed Cases	Active Cases	Critical Cases	Recovered Cases
Mean	4.5404	263.2340	128.91	2.5191	134.60
Maximum	31.000	829.00	538.00	11.000	649.00
Minimum	0.0000	0.0000	0.0000	0.0000	0.0000
Standard Dev.	5.0639	214.6821	124.03	1.9900	128.68
No. of Obs.	235	235	235	235	235

The summary statistics presented in Table 1 for COVID-19 cases in Nigeria indicate daily means of approximately 263, 129, 3, 135, and 5 individuals for confirmed, active, critical, recovered, and death cases, respectively. The corresponding standard deviations are approximately 215, 124, 2, 129, and 5 cases. The notably high standard deviation values compared to the means suggests a significant dispersion from the averages. Additionally, the substantial differences between the maximum and minimum values (ranges) of COVID-19 cases further underscore the wide variability from the corresponding means during the investigated period. From the summary statistics results, it can be concluded that the elevated mean values of confirmed, active, and critical COVID-19 cases in Nigeria over the 235-day period have contributed to a correspondingly high number of COVID-19-related deaths in the studied area.

Omnibus test for COVID-19 cases in Nigeria
The Omnibus Test is a likelihood ratio test that assesses whether the inclusion of all independent variables (confirmed, active, critical, and recovered cases of COVID-19) significantly enhances the model compared to a model with only the intercept (no independent variables added). The results of the Omnibus test are computed and provided in Table 2.

Table 2: Omnibus Test for COVID-19 Cases in Nigeria

Model	LR Chi-Square	Df	P-value
Poisson Regression	1182.160	230	0.000
Negative Binomial Regression	277.010	230	0.000
Generalized Poisson Regression	227.183	230	0.000

The outcomes of the Omnibus test, as indicated in Table 2 for Poisson Regression (PR), Negative Binomial Regression (NBR), and Generalized Poisson Regression (GPR), demonstrate that all predictor variables exhibit statistically significant p-values of 0.000 (i.e., $p < 0.01$). This suggests the overall adequacy and fitness of the models.

Test of multicollinearity for COVID-19 cases in Nigeria
To establish an initial assumption for parameter estimation of COVID-19 cases in Nigeria, a multicollinearity test has been conducted. This test involves calculating variance inflation factor (VIF) values for each predictor variable. Multicollinearity is identified if the VIF value for any independent variable exceeds 10. The multicollinearity test has been performed for each predictor variable, and the results are presented in Table 3.

Table 3: Variance Inflation Factors (VIF) of the Predictor Variables

Predictors	VIF	1/VIF
Confirmed Cases (x_1)	5.6228	0.1778
Active Cases (x_2)	1.6921	0.5910
Critical Cases (x_3)	1.4225	0.7030
Recovered (x_4)	1.2434	0.8042
Mean	2.4952	

The results of the multicollinearity test presented in Table 3 reveal that the variance inflation factor (VIF) values for all predictor variables are below 10. This signifies that all variables are suitable for inclusion in subsequent analyses.

Modeling COVID-19 Cases in Nigeria using Family of Count Data Regression Models
The COVID-19 daily confirmed, active, critical, recovered, and death cases are modeled using a set of count data regression models, specifically: Poisson Regression (PR), Negative Binomial Regression (NBR), and Generalized Poisson Regression (GPR) models and result is presented I Table 4.

Table 4: Parameter Estimates of Count Regression Models

Parameter	Coefficient	Std. Error	z-statistic	$p > |z|$	95% Conf. Interval	
					Lower	Upper
Poisson Regression (PR) Model						
β_0	-2.202	0.8632	-2.551	0.011	-3.894	-0.511
β_1	2.781	0.9775	2.845	0.004	0.865	4.697
β_2	1.007	0.027	37.29	0.000	0.374	3.095
β_3	0.167	0.014	11.93	0.000	0.073	1.571
β_4	-0.775	0.312	-2.484	0.024	-2.663	-0.319
Negative Binomial Regression (NBR) Model						
β_0	-6.513	1.8550	-3.511	0.000	-10.148	-2.877
β_1	5.665	1.9828	2.8571	0.004	1.779	9.552
β_2	1.015	0.0071	142.96	0.000	0.374	3.095
β_3	1.018	0.3111	3.2723	0.001	0.628	3.592
β_4	-0.612	0.254	-2.409	0.037	-2.972	-0.279
Generalized Poisson Regression (GPR) Model						
β_0	-4.521	0.9718	-4.652	0.000	-8.662	-2.285
β_1	3.758	1.0815	3.4748	0.000	1.017	6.218
β_2	2.117	0.1627	13.012	0.000	0.276	4.271
β_3	1.235	0.0995	12.412	0.000	0.384	3.751
β_4	-1.062	0.1243	-8.544	0.000	-3.518	-0.773

The Poisson regression entails modeling the logarithm of the anticipated count in relation to the predictor variables. The parameters of the Poisson regression, as depicted in Table 4, can also be expressed as:

$$y_i = \exp(-2.202 + 2.781 + 1.007 + 0.167 - 0.775)$$

(14)

The intercept signifies the anticipated Poisson regression estimate when all predictor variables in the model are set to zero. In the Poisson regression model reported in the upper panel of Table 4, the intercept exhibits a negative association with COVID-19 related deaths and achieves statistical significance at a 5% level of significance. The logarithm of the expected COVID-19 deaths registers a numerical value below zero (-2.202) when holding all independent variables constant in the model.

The slope coefficients for confirmed (β_1), active (β_2), and critical (β_3) cases of COVID-19 exhibit positive relationships with COVID-19 death cases and achieve statistical significance at the 5% level. This implies that an increase in the numbers of COVID-19 confirmed, active, and critical cases is associated with a corresponding increase in COVID-19-related deaths. Specifically, a one-unit change in confirmed, active, and critical COVID-19 cases is expected to result in changes of 2.781, 1.007, and 0.167 units, respectively, in the differences in the logarithms of expected deaths due to COVID-19, assuming the other predictor variables in the model remain constant.

The slope coefficient for recovered cases (β_4) of COVID-19 displays a negative relationship with COVID-19 death cases and attains statistical significance at the 5% level. This indicates that an increase in the number of COVID-19 recovered cases is associated with a decrease in COVID-19 deaths. To elaborate, a one-unit change in recovered COVID-19 cases is anticipated to result in a reduction of 0.775 units in the difference in the logarithms of expected deaths due to COVID-19, assuming that the other predictor variables in the model remain constant.

In the estimated Negative Binomial regression model presented in the middle panel of Table 4, the intercept demonstrates a negative association with COVID-19 deaths and achieves statistical significance at the 1% level. The logarithm of expected COVID-19 deaths is numerically below zero (-6.513) when all independent variables in the model are held constant.

The slope coefficients for confirmed (β_1), active (β_2), and critical (β_3) cases of COVID-19 exhibit positive relationships with COVID-19 death cases and achieve statistical significance at the 5% level. This implies that an increase in the numbers of COVID-19 confirmed, active, and critical cases is associated with a corresponding increase in COVID-19-related deaths. Specifically, for a one-person increase in the number of confirmed, active, and critical COVID-19 cases, the differences in the logarithms of expected deaths due to COVID-19 are expected to increase by 5.665, 1.015, and 1.018 persons, respectively, assuming the other predictor variables in the model remain constant.

Conversely, the slope coefficient for recovered cases (β_4) of COVID-19 shows a negative relationship with COVID-19 death cases and reaches statistical significance at the 5% level. This indicates that an increase in the number of COVID-19 recovered cases is associated with a decrease in COVID-19 deaths. To be more precise, for a one-person increase in recovered COVID-19 cases, the difference in the logarithms of expected deaths due to COVID-19 is expected to decrease by 0.612 persons, given that the other predictor variables in the model are held constant.

The parameter estimates from the Generalized Poisson Regression model, as presented in the lower panel of Table 4, indicate that all model parameters are statistically significant at the 5% level, as evidenced by p-values numerically less than 0.05. The findings further reveal that the four slope coefficients corresponding to COVID-19 cases are positively associated with COVID-19 death cases, except for the recovered cases. This implies that a one-person increase in the number of confirmed, active, and critical cases of COVID-19 is expected to result in approximately 4 persons, 2 persons, and 1 person increase, respectively, in the number of COVID-19-related deaths in Nigeria.

Additionally, the z-statistics values for confirmed, active, and critical cases (3.4748, 13.012, and 12.412, respectively) are greater than 2. According to the rule of thumb, this suggests that confirmed, active, and critical COVID-19 cases have a positive and significant impact on the number of deaths associated with COVID-19 in Nigeria.

However, the slope coefficient for recovered cases has a negative and significant value (-1.062) at the 5% significance level. This implies that a one-person increase in the number of recovered cases of COVID-19 in Nigeria during the investigated period is expected to lead to approximately a 1-person decrease in the number of deaths due to COVID-19. The z-statistic value for recovered cases is (-8.544), indicating that the recovered cases are less than 2. Following the rule of thumb, this implies that recovered cases have a negative effect on the number of COVID-19-related deaths in Nigeria.

The outcomes of the Poisson regression, Negative Binomial regression, and Generalized Poisson regression models distinctly indicate that the predictor variables positively contributing to the number of COVID-19-related deaths in Nigeria are the confirmed cases (β_1), active cases (β_2), and critical cases (β_3). Conversely, the recovered cases (β_4) exhibit a negative impact on the number of COVID-19-related deaths in Nigeria. This result agrees with the empirical findings of Adams *et al*. (2020).

Goodness of fit test for count data regression models

In Poisson regression modeling, a key assumption is equidispersion, where the mean and variance are equal. However, this assumption is often not met due to the presence of over-dispersion in most cases. Over-dispersion can be identified by examining the ratio of deviance to degrees of freedom (df): a ratio greater than 1 indicates over-dispersion, while a ratio less than 1 suggests under-dispersion. Table 5 presents the results of a goodness-of-fit test used to detect over-dispersion in the three count data regression models.

Table 5: Goodness of Fit Test for Poisson Regression Model

Criterion	Value	Df	Value/df
Poisson Regression Model			
Deviance	257.841	230	1.1210
Pearson Chi-Square	254.982	230	1.1086
Negative Binomial Regression Model			
Deviance	122.332	230	0.5319
Pearson Chi-Square	115.772	230	0.5034
Generalized Poisson Regression Model			
Deviance	120.982	230	0.5260
Pearson Chi-Square	113.879	230	0.4951

The results of the goodness-of-fit test for Poisson regression, outlined in Table 5, reveal that both the deviance/df and Pearson chi-square/df values surpass unity. This indicates an over-dispersion phenomenon in the count of COVID-19 cases in Nigeria. Over-dispersion poses challenges by causing underestimated standard errors and rendering statistical inferences unreliable. To address this issue, Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) models serve as alternative approaches, capable of accommodating the dispersion parameter.

The goodness-of-fit test results for both Negative Binomial regression and Generalized Poisson regression, as reported in Table 5, show that both the deviance/df and Pearson chi-square/df values are below unity. This implies that the modeled count of COVID-19 cases in Nigeria by both models is under-dispersed. Under-dispersion, in contrast to over-dispersion, results in models with unbiased parameter estimates, and it does not impact the standard errors or the inferences drawn from the parameters.

Model comparison using performance criteria

To choose the most suitable model among the three competing count data models for modeling COVID-19 cases in Nigeria, three performance criteria are taken into account: -2log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The preferred model is the one exhibiting the highest -2log-likelihood value and the lowest values for the information criteria. The outcomes of the model comparison are preseted in Table 6.

Table 6: Model Comparison Using Performance Criteria

Model	-2Log-Likelihood	AIC	BIC
PR	-327.905	1021.810	1654.915
NBR	-476.219	1318.438	1951.542
GPR*	**-299.735**	**1006.517**	**1519.361**

Table 6 compares three count data regression models: Poisson Regression (PR), Negative Binomial Regression (NBR), and Generalized Poisson Regression (GPR) in modeling COVID-19 cases in Nigeria. GPR emerges as the preferred model, displaying the highest -2log-likelihood (-299.735), along with the lowest AIC (1006.517) and BIC (1519.361) values. This finding agrees with the previous result of Adams *et al*. (2020), who also found that Generalized Poisson regression was better in modeling COVID-19 data in Nigeria. Beyond these metrics, GPR excels due to its flexibility in handling over- and under-dispersion, accommodating varied variance structures. Its adaptability makes it superior for real-world scenarios where traditional models may falter in capturing data variability, establishing GPR as a robust and versatile choice for count data regression modeling.

CONCLUSION

This study utilized three count data regression models-Poisson, Negative Binomial, and Generalized Poisson-to analyze the factors influencing COVID-19 related deaths in Nigeria, considering confirmed, active, critical, and recovered cases as predictor variables. Secondary data from the Nigeria Centre for Disease Control (NCDC) covering the period from February 29, 2020, to October 19, 2020, was employed. Results indicated that Poisson Regression failed to capture data over-dispersion, leading to the selection of Negative Binomial Regression and Generalized Poisson Regression.

Generalized Poisson Regression emerged as the best-fitting model based on performance criteria such as -2logL, AIC, and BIC. The study identified positive and significant impacts of confirmed, active, and critical cases on COVID-19 deaths, while recovered cases exhibited a negative effect.

The study recommends the use of Generalized Poisson Regression (GPR) for modeling the influence of daily confirmed, critical, active, and recovered COVID-19 cases on related deaths in Nigeria. Additionally, it suggests that the Nigerian Government, particularly through the Presidential Task Force (PTF) on COVID-19 and the Nigeria Centre for Disease Control (NCDC), should prioritize efforts and resources toward monitoring and addressing confirmed, active, and critical cases, as they are identified as the main factors influencing COVID-19 related deaths in the country.

REFERENCES

Adams, S. O., Bamanga, M. A., Olanrewaju, S. O., Yahaya, H. U., and Akano, R. O. (2020). Modeling COVID-19 Cases in Nigeria Using Some Selected Count Data Regression Models. *International Journal of Healthcare and Medical Sciences*, 6(4): 64-73.

Adewole, M. O., Onifade, A. A., Abdullah, F. A., Kasali, F., and Ismail, A. I. M. (2021). Modeling the Dynamics of COVID-19 in Nigeria. *International Journal of Applied Computer and Mathematics*, 7: 67-83.

Agosto, A. and Giudici, P. (2020). A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics. *Risks*, 7: 2-8.

Agresti, A. (2007). An Introduction to Categorical Data Analysis. John Wiley & Sons.

Alzahrani, S. M. (2022). A Log Linear Poisson Autoregressive Model to Understand COVID-19 Dynamics in Saudi Arabia. *Beni-Suef University Journal of Basic and Applied Sciences*, 6(3): 14-27.

Babatunde, A. O. and Igboeli, U. H. (2020). Modeling of COVID-19 Cases of Selected States in Nigeria using Linear and Non-Linear Prediction Models. *Journal of computer science institute*, 17: 390-395.

Barria-Sandoval, C., Ferreira, G., Benz-Parra, K., Lopez-Flores, P. (2021). Prediction of Confirmed Cases of and Deaths Caused by COVID-19 in Chile through Time Series Techniques: A Comparative Study. *PLoS ONE*, 16(4): e0245414.

Cheng, Z. J., Shan, J. (2020). Novel Coronavirus: Where We are and What We Know. Infection 2019. *International Journal of Scientific Research*, **7**(3): 138-149.

Giudici, P., Tarantino, B., and Roy, A. (2022). Bayesian Time-Varying Autoregressive Models of COVID-19 Epidemics. *Biomedical Journal*, 65(1): 212-223.

Hassan, S., Sheikh, F. N., Jamal, S., Ezeh, J. K., Akhtar, A. (2020). Coronavirus (COVID-19): A Review of Clinical Features, Diagnosis, and Treatment. *Cureus*, 12(3): 1-7.

Hilbe, J. M. (2011). Negative Binomial Regression. Cambridge University Press.

Ibrahim, R. R. and Oladipo, H. O. (2020). Forecasting the Spread of COVID-19 in Nigeria using Box-Jenkins Modeling Procedure. *Journal of infectious diseases modeling*, 12(2): 112-125.

Nasution, B. I., Nugraha, Y., Kanggrawan, J. I., and Suherman, A. L. (2021). Forecasting of COVID-19 Cases in Jakarta using Poisson Autoregression. Proceedings of the 9th International Conference on Information and Communication Technology, Yogyakarta, Indonesia, 2021: 112-119.

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

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,* 1: 41-60.

Shereen, M.A., Khan, S., Kazmi, A., Bashir, N., Siddique, R. (2020). COVID-19 Infection: Origin, Transmission and Characteristics of Human Coronaviruses. *Journal of Advanced Research*, 24: 91-98.

Tawiah, K. I., Wahab, A., Asampana, A. K. (2021). Zero-Inflated Time Series Modelling of COVID-19 Deaths in Ghana. *Journal of Environment and Public Health*, 21: 54-69.

WHO, (2020). World Health Organization. Coronavirus Disease 2019 Situation Report=100. Available: https://www.who.int/20200425-sitrep-96-19.

Zhao, J., Han, M., and Wang, Z. (2022). Autoregressive Count Data Modeling on Mobility Patterns to Predict Cases of COVID-19 Infection. *Stochastic Environmental Research and Risk Assessment,* 36: 4185-4200.