International Journal of Statistical Distributions and Applications
2024, Vol. 10, No. 4, pp. 89-100
https://doi.org/10.11648/j.ijsd.20241004.12

Research Article

Performance Assessment of Some Count Data Models to Immunization Coverage Data

Usman Aliyu[1, *], Umar Usman[2], Abubakar Umar Bashar[1], Daha Umar Faruk[3]

[1]Department of Statistics, Waziri Umar Federal Polytechnic, Birnin Kebbi, Nigeria
[2]Department of Statistics, Usmanu Danfodiyo University, Sokoto, Nigeria
[3]Department of Mathematics, Federal University, Birnin Kebbi, Nigeria

Abstract

This research evaluates the performance of various count data models, including Poisson Regression (PR), Zero-Inflated Poisson Regression (ZIP), Zero-Truncated Poisson Regression (ZTP), Truncated Negative Binomial Poisson Regression (TNBP), and Negative Binomial Poisson Regression (NBP), using immunization coverage data from the National Primary Health Care Development Agency (NPHCDA). The study focuses on children under 12 months, assessing model fit using Likelihood Ratio (LR), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) criteria. Analysis conducted with STATA indicates that the Truncated Negative Binomial Poisson Regression (TNBP) ouwtperformed other models in fit and efficiency. Both the ZTeeP and TNBP models demonstrated the best fit, with lower AIC (1959.107) and BIC (2037.649) values and higher Pseudo R-squared values (0.0677 for ZTP and 0.0590 for TNBP), compared to standard models. Age was identified as a significant predictor, negatively associated with immunization status, implying that older infants in the under-12-month category are less likely to receive all vaccinations. The ZTP model showed significant positive effects for antigens such as HepB0, OPV0, BCG, and Measles, with age having a significant negative association. The findings highlight the importance of selecting appropriate statistical models for accurate public health data analysis, enhancing decision-making in immunization programs.

Keywords

PR, ZIP, TPR, NB, NBR, AIC, BIC, LR

1. Introduction

Many experimental situations arise in which we observe the counts of events within a set unit of time, area, volume, length etc. Count data is a statistical data type, a type of data in which the observations can take only the non-negative integer values {0, 1, 2, 3,...}, and where these integers arise from counting, [5]. They are the "realization of a non-negative integer-valued random variable" [4]. As such, the response values take the form of discrete integers [26].

Count data are data that are obtained by counting the number of occurrences of a particular event rather than by taking measurement on some scale [6]. Count data arise in almost every fields of endeavour, including biology, healthcare, psychology, marketing and more. For example, we realize count data from the number of affected persons

*Corresponding author: usmanaliyusta@gmail.com (Usman Aliyu)

Received: 11 July 2024; **Accepted:** 21 October 2024; **Published:** 22 November 2024

with HIV/AIDs; number of death from fatal accident, number of admitted student in our higher institution, number of deaths due to child bearer, number of trade in a time interval, number of given disaster, number of crime on campus per semester and so on [20].

Regression analysis is a statistical tool that describes the functional form of the relationship between the dependent variable (response variable) and one or more independent variables, and produces a statistical model showing the relationship between variables. There are three goals for using regression analysis, first goal is used for estimation, the second goal is used for testing, and the third goal is used for prediction of the dependent variable.

Count data, including zero counts arise in a wide variety of application, hence models for counts have become widely popular in many fields. In statistical field, one may define the count data as that type of observation which takes only the non-negative integers value. Sometimes researchers may count more zeros than the expected. Excess zero can be defined as Zero-Inflation. Excess zero sometimes may be the reason of occurs Over-dispersion (variance a lot larger than mean). Over-dispersion concept is commonly used in the analysis of discrete data. Therefore, linear regression is not applicable procedure to estimate the parameters of predictor due to the asymmetric distribution of the response variable. Under these limitations, Poisson regression and Negative binomial regression are used to model the Count data [2].

1.1. Statement of the Problem

In many fields, count data where observations take non-negative integer values, including zero, are commonly encountered. When modeling such data, a significant challenge arises when the data contains more zero counts than expected, known as zero inflation, leading to over-dispersion. This deviation from the assumptions of standard count data models can negatively impact the accuracy of statistical inferences. Additionally, count data often violates the normality assumption because it is bounded by zero and tends to exhibit skewness.

Previous studies have shown varying performance of count data models in handling these issues. For example, while [14] found the Zero-Inflated Poisson (ZIP) model to outperform the Negative Binomial Poisson model, [10] reported the opposite. Other studies, like [23], found no clear superiority between ZIP and Negative Binomial ZIP models. Such conflicting results highlight the need for further evaluation of these models under different conditions and datasets.

Given these challenges, this research aims to assess the performance of various count data models—specifically Poisson Regression (PR), Zero-Inflated Poisson Regression (ZIP), and Zero-Truncated Poisson Regression (ZTP)—using immunization coverage data. The study will focus on selecting the best model to accurately analyze data from the National Primary Health Care Development Agency (NPHCDA) concerning antigens administered to children under 12 months. By identifying an optimal model, this research seeks to improve the understanding and application of count data modeling in public health contexts, particularly in evaluating immunization coverage.

1.2. Aim and Objectives of Research

The aim of this research is to assess the performance of Poisson regression, Zero-Inflated Poisson Regression and Zero-Truncated Poisson regression analysis on Count data using Simulated and real data (Immunization coverage on antigens administered to children less than 12 months) from the National Primary Health Care Development Agency (NPHCDA).

To achieve this aim the following objectives are formulated:

1) To estimate a suitable Poisson Regression (PR) model, Zero-Inflated Poisson Regression (ZIP), Zero-Truncated Poisson Regression (ZTP), Truncated Negative Binomial Poisson Regression (TNBP) and Negative Binomial Poisson Regression (NBPS) models to the analyse data.
2) To determine which of the models is more efficiency in analyzing count data (Test for efficiency)
3) To determine the best fit model by comparison on LR, AIC and BIC

2. Review of the Existing Research

Abdulkabir, M., *et al*. conducted an empirical study on generalized linear models for count data [1]. They utilized the Poisson regression model and found that its parameters were significant. Testing for over-dispersion using the Quasi-Poisson regression indicated over-dispersion in the Poisson model, leading them to apply the negative binomial regression model. The comparison between the two models, based on the Akaike Information Criterion (AIC), showed the Poisson regression model as the better fit.

Ijomah *et al.* analyzed count data using logistic and Poisson regression models, employing Excel, SPSS 21, and Minitab 16 for their analysis [12]. The study concluded that the logistic regression model provided a superior fit for modeling binary response variables, based on AIC and Bayesian Information Criterion (BIC) values.

Lambert addressed the issue of excess zeros in count data, recommending the zero-inflated Poisson (ZIP) model and applying it to quality control data on manufacturing defects [14]. The study identified two sources of zero counts: a 'perfect state' where no defects could occur, and an 'imperfect state' where defects could still be absent, leading to an increased number of zeros.

Poston and McKibben compared the performance of Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial models in predicting the average number of children ever born to women in the U.S., finding the zero-inflated models superior [21].

Famoye, *et al.* examined count data on road accidents among drivers aged 65 and older, determining that the generalized Poisson regression model was the most suitable for predicting accident counts [8].

Karazsia and Van-Dulmen studied medically attended injuries in children, finding that the zero-inflated Poisson model provided the best fit for the observed data [13].

Zuur, *et al.* discussed various models for count data analysis, including zero-inflated and zero-altered Poisson and negative binomial regression models [27].

Atkins, *et. al.* investigated count outcomes with a skewed distribution and excess zeros, employing a zero-altered Poisson model to analyze alcohol consumption data [3].

Peng compared count data models using injury data from the National Health Interview Survey, concluding that the zero-inflated negative binomial regression model outperformed logistic regression in predicting injury frequency among at-risk populations [19].

Mamun assessed zero-inflated regression models for under-5 mortality data in health research, finding the zero-inflated Poisson and negative binomial models more effective than standard Poisson and negative binomial models [17].

Yang conducted a comparative study of zero-inflated and zero-altered regression models in health surveys, evaluating model performance under different conditions of zero inflation and over-dispersion [24].

Yang *et al.* compared various regression models under different levels of zero inflation and dispersion in health data, demonstrating the superior performance of the zero-inflated and zero-altered negative binomial models [15].

3. Methodology

For the purpose of this research, Raw data will be obtain from National Primary Health Care Development Agency (NPHCDA), Kebbi State on Immunization coverage for antigens administered to children less than 12 months. The antigens are; HepB0, OPV0, BCG, OPV1, PCV1, Penta1, OPV2, PCV2, Penta2, OPV3, PCV3, Penta3, IPV, Measles and Yellow Fever. With children immunization status; Fully Immunized, Partially Immunized and Not Immunized and analysis will be perform using Statistical Software Packages for Windows STATA.

3.1. Poisson Distribution

Poisson distribution is one of the important discrete distributions for Count data in many statistical applications, sometimes called the distribution of rare events. Poisson distribution is often used to account for rare events such as child suicide or ship arrival in the marina, and the use of the Poisson distribution has extended to other fields as communication technologies and statistical quality control. the probability mass function (P.M.F) is given by [5].

$$P(y;\ \mu) = \frac{e^{-\mu}\mu^{y}}{y!} \tag{1}$$

$for\ y = 0, 1, 2, 3, \ldots \ldots ....$

Poisson distribution is specified with a single parameter μ. The parameter (μ) can be a non-integer, the mean and the variance of the Poisson probability mass function are

$E(y) = \mu\ and\ Var\ (y) = \mu$ which is called Equi − dispersion

3.2. Poisson Regression Model (PR)

Poisson regression model is a non-linear (log-linear) regression models and it is convenient for the analysis of count or rate data. Poisson regression is similar to the multiple regression excepting that the response (y) variable is an observed count that follows “the Poisson distribution”. Therefore, the possible values of (y) are “non-negative integers”.

Suppose we have a random sample $x_1, x_2, x_3, \ldots \ldots x_n$ drawn from Poisson distribution, then the P. M. F of x_i, as follows [4].

$$P(x_i;\ \mu_i) = \frac{e^{-\mu_i}\mu_i^{x_i}}{x_i!} \tag{2}$$

$$x_i = 0, 1, 2, 3, \ldots \ldots ....$$

By assumptions of GLM [27], we have

$$Y_i \sim P(\mu_i)$$

$$E(Y_i) = \mu_i\ and\ Var\ (Y_i) = \mu_i$$

$$Log(\mu_i) = X'\beta\ or\ \mu_i = e^{X'\beta}$$

Where

$$X'\beta = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} +, \ldots \ldots + \beta_q X_{iq} \tag{3}$$

$Where\ X_{i1}, X_{i2}, X_{i3} +, \ldots \ldots +$
$X_{iq}\ are\ the\ independent\ variables$

Zero-Inflated Poisson Regression Model (ZIP)

The zero-inflated Poisson regression is used for modelling count data that shows over-dispersion and zero counts (excess zeros)'. This model considers there are two types of data sources, the first source is only zeros count (false zero) and the second source is count variables data (with true zeros) that distributed according to Poisson distribution.

According to [14], the response variable Y_i is independent, and

$$Y_i \sim 0 \text{ with Probability } (\theta_i)$$

$$Y_i \sim Poisson\ (\mu_i) \text{ with Probability } (1 - \theta_i)$$

Where θ_i is the probability that observation (i) is in the always zeros subgroup. Therefore [27],

$$\Pr(Y_i = 0) = \theta_i + (1-\theta_i) \mathrm{X} \Pr(Count\ process\ at\ (i)\ \text{give a zero}) \quad (4)$$

With probability that Y_i is a non-zero count, we have

$$\Pr(Y_i = y_i) = (1-\theta_i) \mathrm{X} \Pr(Count\ process) \quad (5)$$

Furthermore, the probability density function for a ZIP model is given by [19, 7, 14];

$$\Pr(Y_i = y_i) = \begin{cases} \theta_i + (1-\theta_i)e^{-\mu_i}\ if\ y_i = 0 \\ (1-\theta_i)\frac{e^{-\mu_i}{\mu_i}^{x_i}}{x_i!}\ if\ y_i > 0 \end{cases} \quad (6)$$

3.3. Zero-Truncated Poisson Regression Model (ZTP)

Truncated regression models are most commonly used to model zero-truncated count data. Truncations arise when certain values, such as zero are absent from observed data. Truncated distribution arises in cases where the occurrence of an event is limited to values that lie above or below a given threshold, i.e. the Poisson distribution conditioned on being non-zero.

Zero Truncated Poisson (ZTP) regression model is used to model positive count data. When zero count is a potential possible value, but is missing in the data set, we call it zero truncated data. The missing of zero count happens due to the sample scheme, in which the zero count is impossible to be observed [11].

Zero-truncated Poisson (ZTP) regression, introduced by [21], is used to model the always positive counts. (If zero is an admissible value for the dependent variable, then standard Poisson regression is more appropriate [16]). The sampling schemes are most likely the reason that gives rise to the Zero Truncated Poisson (ZTP) model. The density function for Zero-truncated Poisson (ZTP) is expressed (for $y_i = 1, 2, \ldots n$) after the zero value of being truncated, here y_i can be any positive numbers that Y_i takes, the probability for $Y_i = y_i$ is;

$$\Pr(Y_i = y_i |\, y_i > 0) = \frac{\Pr(Y_i = y_i)}{\Pr(Y_i > r)} = \frac{\mu^{y_i} e^{-\mu}}{y_i![1-e^{-\mu}]} = \frac{\mu^{y_i}}{y_i![e^{-\mu}-1]} \quad y_i = 1, 2, \ldots n \quad (7)$$

where n is the number of observation after truncation.

The standard assumption is to use the exponential mean parametrization,

$$\mu_i = \exp({x_i}^T\beta + {z_i}^T\mu_i)\, , i = 1,2,3 \ldots\ldots n \quad (8)$$

In this expression, x_i is a vector of covariates and β is a vector of parameters (fixed effect coefficient). The coefficient β can be interpreted as average proportionate change in the conditional mean $\mathrm{E}(Y_i |\, x_i)$ for a unit change is x_i. Z is a design matrix of random effects clusters and μ is a vector of random effects for that.

3.4. Negative Binomial Poisson Regression (NBPR)

Negative Binomial Poisson Regression (NBPR) is a statistical technique used for modeling count data, particularly when the data exhibit overdispersion. Overdispersion occurs when the variance of the count data is greater than the mean, which violates the assumption of the Poisson distribution that the mean equals the variance. The Negative Binomial distribution, which includes an extra parameter to account for the overdispersion, can address this issue.

Negative Binomial Poisson Regression (NBPR) is an extension of Poisson regression that accounts for overdispersion by introducing an additional parameter to model the variance separately from the mean.

Negative Binomial distribution function is given as

$$\mathrm{g}(y|x) = \frac{\Gamma(y+k)}{\Gamma(u)\,\Gamma(y+1)}\left(\frac{k}{u+k}\right)^k \left(\frac{u}{u+k}\right)^y \quad (9)$$

$where\ y_i = 0, 1, 2, \ldots : k\ and\ n$ are negative binomial parameter with $\mathrm{E}(y) = \mu$ and $\mathrm{Var}(y) = \mu + \mu^2 k$: k mention as disperse parameter which is shown that the data consist over-dispersed

3.5. Truncated Negative Binomial Poisson Regression (TNBP)

Truncated Negative Binomial Poisson Regression (TNBP) is a specialized form of regression used for modeling count data that is truncated at some value. Truncation occurs when the data is not observed or recorded below or above a certain threshold. This can happen in various scenarios, such as when data collection processes exclude certain counts or when counts naturally cannot be below or above a certain number.

3.6. Model Selection

It is important that we have one or more a criterion to consider the best results and choose the appropriate model for data representation. 'There are several methods that provide a measure for selecting the appropriate model'. So, we will use the following methods to selecting best model.

3.7. Akaike Information Criterion

The Akaike information criterion (AIC) is an evaluating model fit for a given data among different types of non-nested models. It is widely used for statistical inference, and its formula is given as

$$AIC = -2logL + 2k \quad (10)$$

Where

L: The maximum likelihood function of the model.

K: Number of model parameters.

The model with minimum AIC value is chosen as the best model to fit the data.

Bayesian Information Criterion

The Bayesian information criterion (BIC) is another estimator for evaluating model fit for a given data among different types of non-nested models, and its formula is given as

$$BIC = -2logL + klogn \quad (11)$$

Where

L: The maximum likelihood function of the model.

k: Number of model parameters.

n: The number of observations, or the sample size.

The model with minimum BIC value is chosen as the best model to fit the data.

The Likelihood Ratio Test

The likelihood ratio test (LR) is a statistical test used to compare two "nested models" and determine which model fits the data better, its formula is given as

$$LR = -2log\frac{L_1}{L_2} \quad (12)$$

Where

L_1: The likelihood of the first model.

L_2: The likelihood of the second model.

4. Analysis and Results

Raw data was obtained from National Primary Health Care Development Agency (NPHCDA), Kebbi State on Immunization coverage for antigens administered to children less than 12 months. For convenience, the following variables were derived that is coded as:

The antigens are; HepB0 (1 = Yes, 0= No), OPV0 (1 = Yes, 0= No), BCG (1 = Yes, 0= No), OPV1, PCV1 (1 = Yes, 0= No), Penta1 (1 = Yes, 0= No), OPV2 (1 = Yes, 0= No), PCV2 (1 = Yes, 0= No), Penta2 (1 = Yes, 0= No), OPV3 (1 = Yes, 0= No), PCV3 (1 = Yes, 0= No), Penta3 (1 = Yes, 0= No), IPV (1 = Yes, 0= No), Measles (1 = Yes, 0= No), Yellow Fever (1 = Yes, 0= No), Sex/Gender (1 = male, 0 = female) and Ages (count are considered) i.e all the independent variables were coded to 0 and 1 except for age which was considered as numeric variable with children immunization status; Fully Immunized, Partially Immunized and Not Immunized. The collected data was analyzed using statistical software package STATA and the following results were obtained:

Table 1. *Poisson regression.*

				Number of obs	=	750
				LR chi2(16)	=	74.59
				Prob > chi2	=	0.0000
Log likelihood = -1060.4086				Pseudo R2	=	0.0340
Status	Coef.	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age	-.0107634	.0079186	-1.36	0.174	-.0262835	.0047568
Gender	.0065227	.0477234	0.14	0.891	-.0870134	.1000589
HepB0	.2401427	.1041869	2.30	0.021	.0359401	.4443453
OPV0	.2851453	.1329035	2.15	0.032	.0246593	.5456313
BCG	.2762152	.103503	2.67	0.008	.073353	.4790773
PCV1	-.0028828	.1051621	-0.03	0.978	-.2089967	.2032311
Penta1	.0900432	.102993	0.87	0.382	-.1118194	.2919058
OPV2	.048612	.1030479	0.47	0.637	-.1533581	.2505822
PCV2	-.0238464	.1144921	-0.21	0.835	-.2482468	.2005541
Penta2	.020943	.1203517	0.17	0.862	-.214942	.256828
OPV3	.0835448	.1886173	0.44	0.658	-.2861383	.453228
PCV3	-.0717459	.1185837	-0.61	0.545	-.3041656	.1606739
Penta3	.0190417	.1969425	0.10	0.923	-.3669585	.405042

IPV	-.0999678	.2296027	-0.44	0.663	-.5499809	.3500452
Measles	.4534085	.3824454	1.19	0.236	-.2961708	1.202988
Yellow Fever	-.2064341	.2748705	-0.75	0.453	-.7451704	.3323022
_cons	.240673	.1655188	1.45	0.146	-.083738	.565084

. estat ic

Table 2. Truncated Poisson regression.

				Number of obs	=	750
Truncation point: 0				LR chi2(16)	=	139.84
				Prob > chi2	=	0.0000
Log likelihood = -962.55365				Pseudo R2	=	0.0677
Status	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Age	-.0305948	.0107627	-2.84	0.004	-.0516892	-.0095003
Gender	.0059933	.0559341	0.11	0.915	-.1036355	.115622
HepB0	.3743356	.1295592	2.89	0.004	.1204042	.628267
OPV0	.5417303	.1853856	2.92	0.003	.1783812	.9050793
BCG	.501562	.1432965	3.50	0.000	.2207059	.7824181
PCV1	.0608188	.1313071	0.46	0.643	-.1965383	.318176
Penta1	.1377057	.1246479	1.10	0.269	-.1065996	.382011
OPV2	.0857016	.1269404	0.68	0.500	-.1630971	.3345002
PCV2	-.0301775	.1413009	-0.21	0.831	-.3071221	.2467671
Penta2	.0680832	.1497724	0.45	0.649	-.2254652	.3616317
OPV3	.1550095	.2406803	0.64	0.520	-.3167151	.6267342
PCV3	-.1305315	.1497319	-0.87	0.383	-.4240005	.1629376
Penta3	.0772947	.2519963	0.31	0.759	-.416609	.5711983
IPV	-.2428938	.2999722	-0.81	0.418	-.8308286	.345041
Measles	1.380408	.5430408	2.54	0.011	.3160674	2.444748
Yellow Fever	-.59394	.373232	-1.59	0.112	-1.325461	.1375813
_cons	-.4113474	.233137	-1.76	0.078	-.8682874	.0455927

. estat ic

Table 3. Truncated negative binomial regression.

				Number of obs	=	750
Truncation point: 0				LR chi2(15)	=	120.72
Dispersion= mean				Prob > chi2	=	0.0000
Log likelihood = -962.55365				Pseudo R2	=	0.0590
Status	Coef.	Std. Err.	Z	P>|z|	[95% Conf.	Interval]
Age	-.0305948	.0107627	-2.84	0.004	-.0516892	-.0095003

Gender	.0059933	.0559341	0.11	0.915	-.1036355	.115622
HepB0	.3743356	.1295592	2.89	0.004	.1204042	.6282669
OPV0	.5417302	.1853855	2.92	0.003	.1783812	.9050792
BCG	.5015619	.1432965	3.50	0.000	.2207059	.782418
PCV1	.0608188	.1313071	0.46	0.643	-.1965383	.3181759
Penta1	.1377057	.1246478	1.10	0.269	-.1065996	.3820109
OPV2	.0857016	.1269404	0.68	0.500	-.1630971	.3345002
PCV2	-.0301775	.1413008	-0.21	0.831	-.3071221	.2467671
Penta2	.0680832	.1497724	0.45	0.649	-.2254652	.3616316
OPV3	.1550095	.2406802	0.64	0.520	-.3167151	.6267341
PCV3	-.1305314	.1497319	-0.87	0.383	-.4240005	.1629376
Penta3	.0772946	.2519963	0.31	0.759	-.416609	.5711982
IPV	-.2428937	.2999722	-0.81	0.418	-.8308285	.345041
Measles	1.380408	.5430407	2.54	0.011	.3160673	2.444748
Yellow Fever	-.5939399	.373232	-1.59	0.112	-1.325461	.1375813
_cons	-.4113472	.233137	-1.76	0.078	-.8682872	.0455928
/lnalpha	-23.13146	.	.	.		
Alpha	9.00e-11	.	.	.		

Likelihood-ratio test of alpha=0: chibar2(01) = 0.00 Prob>=chibar2 = 1.000

. estat ic

Table 4. *Negative binomial regression.*

				Number of obs	=	750
				LR chi2(16)	=	74.59
Dispersion= mean				Prob > chi2	=	0.0000
Log likelihood = -1060.4086				Pseudo R2	=	0.0340
Status	Coef.	Std. Err.	Z	P>|z|	[95% Conf.	Interval]
Age	-.0107634	.0079186	-1.36	0.174	-.0262835	.0047568
Gender	.0065227	.0477234	0.14	0.891	-.0870134	.1000589
HepB0	.2401427	.1041869	2.30	0.021	.0359401	.4443453
OPV0	.2851453	.1329035	2.15	0.032	.0246593	.5456313
BCG	.2762152	.103503	2.67	0.008	.073353	.4790773
PCV1	-.0028828	.1051621	-0.03	0.978	-.2089967	.2032311
Penta1	.0900432	.102993	0.87	0.382	-.1118194	.2919058
OPV2	.048612	.1030479	0.47	0.637	-.1533581	.2505822
PCV2	-.0238464	.1144921	-0.21	0.835	-.2482468	.2005541
Penta2	.020943	.1203517	0.17	0.862	-.214942	.256828
OPV3	.0835448	.1886173	0.44	0.658	-.2861383	.453228
PCV3	-.0717459	.1185837	-0.61	0.545	-.3041656	.1606739

Penta3	.0190417	.1969425	0.10	0.923	-.3669585	.405042
IPV	-.0999678	.2296027	-0.44	0.663	-.5499809	.3500452
Measles	.4534085	.3824454	1.19	0.236	-.2961707	1.202988
Yellow Fever	-.2064341	.2748705	-0.75	0.453	-.7451704	.3323021
_cons	.240673	.1655188	1.45	0.146	-.0837379	.565084
/lnalpha	-38.46945	.	.			
Alpha	1.96e-17	.	.	.		

Likelihood-ratio test of alpha=0: chibar2(01) = 0.00 Prob>=chibar2 = 1.000
. estat ic

Table 5. *Zero-inflated Poisson regression.*

				Number of obs	=	750
				Nonzero obs	=	723
				Zero obs	=	27
Inflation model = logit				LR chi2(16)	=	74.58
Log likelihood = -866.1446				Prob > chi2	=	0.0000
Status	Coef.	Std. Err.	Z	P>|z|	[95% Conf.	Interval]
Age	-.0247416	.0136018	-1.82	0.069	-.0514006	.0019174
Gender	.0089983	.0617592	0.15	0.884	-.1120475	.1300441
HepB0	.4217198	.1411287	2.99	0.003	.1451126	.698327
OPV0	.4870499	.1955768	2.49	0.013	.1037265	.8703734
BCG	.4826598	.1535231	3.14	0.002	.18176	.7835596
PCV1	.0171423	.1439135	0.12	0.905	-.264923	.2992077
Penta1	.1569799	.1357069	1.16	0.247	-.1090008	.4229605
OPV2	.0922848	.1377348	0.67	0.503	-.1776704	.36224
PCV2	-.0306114	.1533053	-0.20	0.842	-.3310842	.2698614
Penta2	.0386251	.1624761	0.24	0.812	-.2798222	.3570723
OPV3	.1593018	.2586079	0.62	0.538	-.3475604	.666164
PCV3	-.12049	.1625509	-0.74	0.459	-.4390838	.1981038
Penta3	.0430333	.2708739	0.16	0.874	-.4878697	.5739364
IPV	-.1703819	.3256836	-0.52	0.601	-.80871	.4679463
Measles	.9598294	.6506126	1.48	0.140	-.3153478	2.235007
Yellow Fever	-.3935698	.4134645	-0.95	0.341	-1.203945	.4168056
_cons	-.7665347	.259896	-2.95	0.003	-1.275921	-.257148
Inflate						
Status	-55.17046	1713445	-0.00	1.000	-3358345	3358234
_cons	30.59125	1713403	0.00	1.000	-3358178	3358239

Table 6. Model Comparison.

Model	Obs	df	Log likelihood	Likelihood Ratio Test	AIC	BIC
Poisson regression	750	17	-1060.409	74.59	2154.817	2233.358
Truncated Poisson regression	750	17	-962.5537	139.84	1959.107	2037.649
Truncated negative binomial regression	750	17	-962.5537	120.72	1959.107	2037.649
Negative binomial regression	750	17	-1060.409	74.59	2154.817	2233.358
Zero inflated Poisson regression	750	19	-866.1446	74.58	1770.289	1858.071

5. Discussion and Interpretation of Results

Results obtained from the Poisson Regression (PR) Model with Log likelihood = -1060.4086 LR chi2(16) = 74.59, Prob > chi2 = 0.0000, Pseudo R2 = 0.0340 have coefficients and Significance HepB0: Coefficient = 0.2401, p-value = 0.021 (significant), OPV0: Coefficient = 0.2851, p-value = 0.032 (significant), BCG: Coefficient = 0.2762, p-value = 0.008 (significant). These significant variables suggest that HepB0, OPV0, and BCG vaccinations positively impact the status variable (Immunization coverage) while other variables (Age, Gender, PCV1, Penta1, OPV2, PCV2, Penta2, OPV3, PCV3, Penta3, IPV, Measles, Yellow Fever) are not statistically significant (p > 0.05). The model Fit for Poisson Regression: Log likelihood: -1060.4086, AIC: 2154.817 and BIC: 2233.358.

Zero-Truncated Poisson Regression (ZTP) Model with log likelihood = -962.55365, LR chi2(16) = 139.84, Prob > chi2 = 0.0000, Pseudo R2 = 0.0677, with Coefficients and Significance: Age: Coefficient = -0.0306, p-value = 0.004 (significant), HepB0: Coefficient = 0.3743, p-value = 0.004 (significant), OPV0: Coefficient = 0.5417, p-value = 0.003 (significant), BCG: Coefficient = 0.5016, p-value = 0.000 (significant), Measles: Coefficient = 1.3804, p-value = 0.011 (significant). Here, Age is significant and negatively associated with the status, indicating that as age increases, the count of immunization status decreases. Other significant variables positively affect the status (Immunization coverage) while other variables are not statistically significant (p > 0.05). The model fit: Log likelihood: -962.55365, AIC: 1959.107 and BIC: 2037.649.

Truncated Negative Binomial Poisson Regression (TNBP) Model with log likelihood = -962.55365, LR chi2(15) = 120.72, Prob > chi2 = 0.0000, Pseudo R2 = 0.0590, with Coefficients and Significance: Age: Coefficient = -0.0306, p-value = 0.004 (significant), HepB0: Coefficient = 0.3743, p-value = 0.004 (significant), OPV0: Coefficient = 0.5417, p-value = 0.003 (significant), BCG: Coefficient = 0.5016, p-value = 0.000 (significant), Measles: Coefficient = 1.3804, p-value = 0.011 (significant). The TNBP model has similar significant variables as the ZTP model, suggesting robustness in results across these models.

While other variables are not statistically significant (p > 0.05). The model Fit: Log likelihood: -962.55365, AIC: 1959.107 and BIC: 2037.649.

Negative Binomial Poisson Regression Model with Log likelihood = -1060.4086, LR chi2(16) = 74.59 Prob > chi2 = 0.0000, Pseudo R2 = 0.0340, with Coefficients and Significance: HepB0: Coefficient = 0.2401, p-value = 0.021 (significant), OPV0: Coefficient = 0.2851, p-value = 0.032 (significant), BCG: Coefficient = 0.2762, p-value = 0.008 (significant), The NBP model has the same significant variables as the PR model. Other variables are not statistically significant (p > 0.05). The Model Fit: Log likelihood: -1060.4086, AIC: 2154.817 and BIC: 2233.358.

5.1. Model Comparison

Poisson Regression: AIC: 2154.817 BIC: 2233.358 Likelihood: -1060.4086

Zero-Truncated Poisson Regression: AIC: 1959.107 BIC: 2037.649

Truncated Negative Binomial Poisson Regression: AIC: 1959.107 BIC: 2037.649

Negative Binomial Poisson Regression: AIC: 2154.817 BIC: 2233.358

5.2. Efficiency Test

To determine the most efficient model, compare the AIC and BIC values. Lower values indicate a better fit.

Best Fit Models: ZTP and TNBP have the lowest AIC and BIC values, indicating they fit the data better than the Poisson and Negative Binomial models.

To determine which method is more efficient in analyzing count data.

AIC/BIC Comparison: Lower AIC and BIC values indicate better model fit. The Truncated Poisson and Truncated Negative Binomial models have the lowest AIC (1959.107) and BIC (2037.649) values, suggesting better performance than the Poisson and Negative Binomial models.

Pseudo R2: Higher values indicate better model fit. The Truncated Poisson (0.0677) and Truncated Negative Binomial (0.0590) models have higher Pseudo R2 values compared to the Poisson (0.0340) and Negative Binomial (0.0340) models.

It can be clearly seen from the results that the best models: Zero-Truncated Poisson Regression (ZTP) and Truncated Negative Binomial Poisson Regression (TNBP). These results suggest that the ZTP and TNBP models provide a more efficient analysis of count data on immunization coverage, with significant predictors being crucial for understanding factors influencing immunization status in children under 12 months.

Significant Predictors: HepB0, OPV0, BCG, Measles (positive), Age (negative) in ZTP and TNBP models. HepB0, OPV0, and BCG consistently show significant positive effects on immunization status across all models, indicating these vaccines are important predictors of immunization coverage. Age is negatively associated with immunization status in the truncated models, implying that older children within the <12 months category might be less likely to receive all immunizations. Measles vaccination is significant in the truncated models, suggesting an impactful role in immunization status.

6. Major Findings

1) Significant Predictors:

Across all models, HepB0, OPV0, and BCG are consistently significant predictors of the immunization status, indicating their strong influence on the count of immunizations administered to children under 12 months.

The Age variable is only significant in the truncated models (TPR and TNBP), suggesting that older children within the sample have a lower count of immunizations.

The Measles variable is significant in the truncated models but not in the standard Poisson or Negative Binomial models.

2) Model Fit and Efficiency:

The Truncated Poisson and Truncated Negative Binomial models show better fit (higher log likelihood) compared to the standard Poisson and Negative Binomial models, suggesting that accounting for the truncation (i.e., excluding zero counts) provides a more accurate model for this dataset.

The AIC and BIC values also indicate better model fit for the truncated models compared to the non-truncated models.

3) Efficiency of Models:

Based on the log likelihood, AIC, and BIC, the Truncated Poisson Regression (TPR) appears to be the most efficient model for analyzing the count data in this study. It provides the best fit among the models tested and identifies significant predictors more effectively.

7. Conclusion

Base on the analysis as well as the results obtained, the following are conclusion reach:

1) Age has the highest impact on diabetes followed by gender
2) HepB0, OPV0, and BCG are significant predictors of immunization coverage.
3) The Zero-Truncated Poisson Regression (ZTP) and Truncated Negative Binomial Poisson Regression (TNBP) models are identified as the most efficient model for this dataset, providing the best fit and identifying significant predictors effectively and more efficient for analyzing count data on immunization coverage, based on their lower AIC and BIC values and higher Pseudo R2. These models provide a better fit and capture the significant predictors of immunization status more effectively, suggesting they provide a better fit to the data compared to the Poisson Regression (PR) and Negative Binomial Poisson Regression (NBP) models.
4) Both ZTP and TNBP models indicate significant effects for Age, HepB0, OPV0, BCG, and Measles.
5) Among these models, the Truncated Negative Binomial Poisson Regression (TNBP) can handle overdispersion better than the ZTP model, which might be preferable for count data with overdispersion.
6) Therefore, the TNBP model is recommended for analyzing the count data on immunization coverage among children under 12 months.

8. Recommendations

In the light of the above it is recommended that:

1) Future research could focus on further exploring the effects of age and measles immunization, as they showed significance in the truncated models.
2) These results offer valuable insights for the National Primary Health Care Development Agency (NPHCDA) to improve immunization strategies and ensure higher coverage for children under 12 months.
3) Case study and other agency or organization concern should use the results of this research as a means of justification concerning cases of immunization coverage in kebbi state.
4) Health personnel should intensify efforts in creating awareness on the importance of immunization coverage in the state through village heads, local chiefs, imams and pastors, town announcers, social gatherings and mass media.
5) The Ministry of Health (MOH) could develop a strategic plan to build poly-clinics in every district capital and cheap-compound health facilities, at least in every community. This could help to access health care.
6) Further research could be done in this area by considering other cases from the clinics in and around the Birnin Kebbi metropolis in order to examine the spatial variation. This would even improve upon the scope.

Abbreviations

PR	Poisson Regression
ZIP	Zero-Inflated Poisson Regression
ZTP	Zero-Truncated Poisson Regression
TNBP	Truncated Negative Binomial Poisson Regression
NBPR	Negative Binomial Poisson Regression
NPHCDA	The National Primary Health Care Development Agency
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
LR	Likelihood Ratio Test

Conflicts of Interest

The Authors declare no conflict of interest.

References

[1] Abdulkabir, M., Udokang, A. E., Raji, S. T., and Bello L. K. (2015), An empirical study of generalized linear model for count data Applied & Computational Mathematics 2015, 4:5 http://dx.doi.org/10.4172/2168-9679.1000253

[2] Agresti, A. (2007), An Introduction to Categorical Data Analysis, Second Edition, Wiley, Inc., New York.

[3] Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J. and Neighbors, C., (2013). "A tutorial on count regression and zero-altered count models for longitudinal substance use data": Correction to Atkins *et al.* (2012). Psychology of Addictive Behaviors, 27(2), 379.

[4] Cameron, A. C., and Trivedi, P. K. (1998). Regression analysis of count data. New York: Cambridge University Press.

[5] Cameron, A. C.; Trivedi, P. K. (2013). Regression Analysis of Count Data Book (Second ed.). Cambridge University Press. ISBN 978-1-107-66727-3.

[6] Everitt B. S. (2002): The Cambridge Dictionary of Statistics (2nd ed.). New York: Cambridge University Press.

[7] Famoye, F. and Singh, K. P., (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, *4*(1), pp. 117-130.

[8] Famoye, F., Wulu, J. T. and Singh, K. P., (2004). On the generalized Poisson regression model with an application to accident data. Journal of Data Science, 2(2004), pp. 287-295.

[9] Greene, C. B. (1994). Testing for over dispersion in Poisson and binomial regression models. Journal of the American Statistical Association 87(418): 451-57.

[10] Greene, W. H. (2003), Econometric Analysis; New York University, Upper Saddle River, New Jersey, Fifth Edition.

[11] Hardin J. W., and Hilbe, J. M. (2007) Generalized Linear Models and Extensions, Second Edition. 2nd ed. Stata Press; 2007.

[12] Ijomah, M. A., Biu, E. O., and Mgbeahurike, C. (2018) Assessing Logistic and Poisson Regression Model in Analyzing Count Data. International Journal of Applied Science and Mathematical Theory ISSN 2489-009X Vol. 4 No. 1 2018. www.iiardpub.org

[13] Karazsia, B. T. and Van Dulmen, M. H., (2008). Regression models for count data: Illustrations using longitudinal predictors of childhood injury. Journal of pediatric psychology, 33(10), pp. 1076-1084.

[14] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34(1), pp. 1-14.

[15] Yang, S., Harlow, L. I., Puggioni, G., & Redding, C. A. (2017). A comparison of different methods of zero-inflated data analysis and an application in health surveys. Journal of Modern Applied Statistical Methods, 16(1), 518-543. http://dx.doi.org/10.22237/jmasm/1493598600

[16] Long, J. S., and Freese, J. (2005) Regression Models for Categorical Dependent Variables Using Stata, Second Edition. 2nd ed. Stata Press; 2005.

[17] Mamun, M. A. A., (2014). Zero-inflated regression models for count data: an application to under-5 deaths (Master thesis, Ball State University) Muncie, Indiana.

[18] Mwalili, S. M., Lesaffre, E. and Declerck, D., (2008). The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical methods in medical research*, *17*(2), pp. 123-139.

[19] Peng, J. (2013). Count Data Models for Injury Data from the National Health Interview Survey (NHIS) (Doctoral dissertation, The Ohio State University).

[20] Piza, E. L. (2012): Using Poisson and Negative Binomial Regression Models to Measure the Influence of *R*isk on Crime incident Counts. *R*utgers Center on Public Security. Riskterrainmodeling.com

[21] Poston Jr, D. L. and McKibben, S. L., (2003). Using zero-inflated count regression models to estimate the fertility of US women. Journal of Modern Applied Statistical Methods, 2(2), p. 10.

[22] Shaw, D. (1988), 'On-site samples' regression problems of non-negative integers, truncation, and endogenous stratification', Journal of Econometrics, 37, 211-223; 2005.

[23] Slymen, D., Ayala, G., Arredondo, E., and Elder, J., (2006), A demonstration of modeling count data with an application to physical activity. Epidemiologic Perspectives and Innovations 3: 1-9.

[24] Yang, S., (2014). A comparison of different methods of zero-inflated data analysis and its application in health surveys (Master thesis, Rhode Island University).

[25] Yang, S., Harlow, L. L., Puggioni, G. and Redding, C. A., 2017. A Comparison of Different Methods of Zero-Inflated Data Analysis and an Application in Health Surveys. Journal of Modern Applied Statistical Methods, 16(1), p. 29.

[26] Zorn, C. J. W. (1996). Evaluating zero-inflated and hurdle Poisson specifications. Midwest Political Science Association, 1-16.

[27] Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A. and Smith, G. M., (2009). Mixed effects models and extensions in ecology with R. Springer Science & Business Media.