A Poisson Regression Approach in Modeling the Prevalence Rate of HIV/AIDS in Ibadan, Oyo State, Nigeria

[1]Odusanya, O. Ayodeji, [2]Adedotun, A. Funmi, [3]Obadina, O. Gabriel & [4]Adebukunola, O. Solomon
[1&4].S Adegbenro ICT Polytechnic, Itori-Ewekoro, Ogun State, Nigeria
[2&3]Olabisi Onabanjo University, Ago-Iwoye, Nigeria
E-mails: Oloriebiodus@gmail.com, adedayofunmmy@yahoo.com, olugbenga.obadina@yahoo.com, gbengasolomon1982@gmail.com
Phones: +2348062739675, +2347063518526, +2348057235295, +2347038581606

ABSTRACT

The study was carried out to model the occurrences of rare infectious disease. Rare infectious diseases are those diseases that are not common but highly communicable and deadly which follow Poisson distribution and are mostly not stationary in nature using Poisson Regression. The data on HIV/AIDS were collected from State Hospital, Ibadan, Oyo State. The estimation of Poisson regression parameters was based on Ordinary Least Squares method and Maximum Likelihood Estimation and the test of hypothesis following t-distribution and F-distribution. The goodness of fit of the model was based on R^2. The fitted model AIDS prevalence rate and the result showed that best R^2 was obtained for the Maximum Likelihood Estimation with a value of 0.998 and The fitted model is given as $In(mean\ case) = 5.246723 + 0.047856female - 0.644854male - 0.113363married - 0.0607single - 0.12346quarter1 - 0.148015quarter2 - 0.083806quarter3 - 0.239595quarter4 - 0.039282mode\ of\ Transmission$, with R^2 value of 0.98, which implied that about 99.8% of the total variation in the cases of occurrence is jointly explained by the age, marital status, time and mode of transmission. In conclusion, this study provides a framework for the investigation of other rare infectious diseases in future.

Keywords:Poisson Regression, Modeling, Prevalence Rate, HIV/AIDS, Ibadan, Oyo State, Nigeria

iSTEAMS Proceedings Reference Format
Odusanya, O. Ayodeji, Adedotun, A. Funmi, Obadina, O. Gabriel & Adebukunola, O. Solomon (2019): A Poisson Regression Approach in Modeling the Prevalence Rate of HIV/AIDS in Ibadan, Oyo State, Nigeria. Proceedings of the 17th iSTEAMS Multidisciplinary Research Conference, D.S. Adegbenro ICT Polytechnic, Itori-Ewekoro, Ogun State, Nigeria, 21st – 23rd July, 2019. Pp 259-274. www.isteams.net - DOI Affix - https://doi.org/ 10.22624/AIMS/iSTEAMS-2019/V17N1P25

1. INTRODUCTION

Mortality rate among AIDS infection should be given full intention in Nigeria since the World Health Organization (WHO) in recent years worldwide recorded the highest figure about this problem. There are several factors that affected mortality AIDS infection patients. Regression technique was used to estimate the relationship between exposure and death rate. The usual regression method models, widely used for mortality data in environment epidemiology is the Poisson regression modeling (McCullagh and Nelder, 2009). While existing over dispersion is a common problem with Poisson regression when conditional variance is greater than conditional mean in the observed count data. When absence of over dispersion in Poisson regression, negative binomial has been proven able to some situations when the Poisson model is poor fit (Cameron and Trevedi, 2008). This study considered a standardization of death rate via negative binomial and Poisson regression to compare of the models.

Thus, there are several methods considered a new approach in environmental epidemiology such as generalized linear mixed effect model was used the penalized splinnes as the smoothing methods (Chunag et al., 2007) or generalized additive model that use natural cubic splines as the smoothing method (Morgan.et al., 2008) with Poisson and negative binomial regression.

No infection in recent times has been so feared like HIV/AIDS. AIDS is the end stage of infections with Human Immuno Deficiency Virus (HIV) characterized by a cluster of illness (WHO, 2004). The first case of Acquired Immune Deficiency Syndrome (AIDS) known in Nigeria was in sexually active thirteen-year-old girl in 2004 (FMOHHSS, 2002) and since then it has maintained a metric system like a mobile network. Evian (2000) stated that, HIV/AIDS epidemic spread has been in a rapacious speed. World Health Organization (WHO, 2003), noted that the percentage of the global population infected with HIV/AIDS is situated in sub-Saharan Africa; while the United Nations programme on HIV/AIDS (UNAIDS, 2003) put the prevalence rate in Nigeria at about 5.6 per cent and two hundred thousand deaths. This may not augur well for the future of Nigeria if the groups at risk (the adolescents) are not taught its prevention and control. Ichoku (2003) also pointed out that, due to culture of silence, secrecy and pretention attitude to HIV/AIDS and AIDS patients, many Nigerians carry the disease and spread it knowingly or unknowingly to other people leading to increase in the number of HIV infections and deaths through AIDS. Those cases of HIV/AIDS may not be said to be absent in the Ibadan, Oyo State, where adolescents are a part of the population.

The adolescent being sexually active, need to be provided with adequate information and messages on AIDS prevention and control, since they are most likely to contract the disease and spread it rapidly. In Nigeria, even though the exact number of people living with HIV/AIDS is not known, the Federal Ministry of Health had estimated that over two million Nigerians were infected with HIV by mid-2009. This number may be a mere speculation given the fact that it is practically difficult to give accurate information on seroprevalence rate in Nigeria because like other developing countries, under-reporting the result of management and logistics problems can hamper efficient data collection and analysis (Momodu and Momodu, 2008).

The primary aim of this study is to investigate and model the prevalence of HIV/AIDS in Ibadan, Oyo State, Nigeria. In other to achieve this, the following were put into consideration: Understanding the demographic analysis of the prevalence of the incidence of HIV/AIDS, modelling the prevalence using Poisson Regression and recommending to stakeholders on the prevalence and preferring solutions. Parodi and Bottarelli (2006) described Poisson regression model as a technique used to describe count data which is a function of a set of predictor variables. In the last two decades it has been extensively used both in human and in veterinary Epidemiology to investigate the incidence and mortality of chronic diseases. Among its numerous applications, Poisson regression has been mainly applied to compare exposed and unexposed cohorts and to evaluate the clinical course of ill subject. This review provides a description of the Poisson regression in the framework of the prospective cohort study, which represents the conceptual ground of most epidemiological investigations.

2. METHODOLOGY

Generalized Linear Models (GLM)

Generalized linear models (GLM) was first introduced by Nelder and Wedderburn (1972). They provided a unified framework to study various regression models, rather than a separate study for each individual regression. Generalized linear models (GLM) are extensions of classical linear models. It includes linear regression models, analysis of variance models, logistic regression models, Poisson regression models, loglinear models, as well as many other models. The above models share a number of unique properties, such as linearity and a common method for parameter estimation. A generalized linear model consists of three components:

1. A random component, specifying the conditional distribution of the response variable, Y_i given the explanatory variables.

2. A linear function of the regressors, called the linear predictor,

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = x_i\beta \qquad (1)$$

on which the expected value μ_i of Y_i depends.

3. An invertible link function, $g(\mu_i) = \eta_i$ which transforms the expectation of the response to the linear predictor.

The inverse of the link function is sometimes called the mean function:

$$g^{-1}(\eta_i) = \mu_i \qquad (2)$$

For traditional linear models in which the random component consists of the assumption that the response variable follows the Normal distribution, the canonical link function is the identity link. The identity link specifies that the expected mean of the response variable is identical to the linear predictor, rather than to a non-linear function of the linear predictor. The Generalized Linear Model is an extension of the General Linear Model to include response variables that follow any probability distribution in the exponential family of distributions. The exponential family includes such useful distributions as the Normal, Binomial, Poisson, Multinomial, Gamma, Negative Binomial, and others.

The Poisson Distribution

The Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume. The Poisson regression model is a technique used to describe count data as a function of a set of predictor variables. In the last two decades it has been extensively used both in human and in veterinary epidemiology to investigate the incidence and mortality of chronic diseases. Among its numerous applications, Poisson regression has been mainly applied to compare exposed and unexposed cohorts and to evaluate the clinical course of ill subjects.

The distribution was first introduced by Simeon-Denis Poisson (1781–1840) and published together with his probability theory, in 1838 in his work Recherchessur la probabilite des jugements en matierecriminelleetenmatierecivile ("Research on the Probability of Judgments in Criminal and Civil Matters"). The work focused on certain random variables N that count, among other things, the number of discrete occurrences (sometimes called "arrivals") that take place during a time-interval of given length.

If the expected number of occurrences in this interval is λ, then the probability that there are exactly k occurrences (k being a non-negative integer, k = 0, 1, 2…) is equal to

$$f(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (3)$$

Where;

- e is the base of the natural logarithm (e = 2.71828...)
- k is the number of occurrences of an event - the probability of which is given by the function
- k! is the factorial of k
- λ is a positive real number, equal to the expected number of occurrences that occur during the given interval. For instance, if the events occur on average 4 times per minute, and one is interested in probability for k times of events occurring in a 10-minute interval, one would use as the model a Poisson distribution with λ= 10×4 = 40.

The parameter λ is not only the mean number of occurrences, k but also its variance

$$\sigma_k^2 = E(k^2) - [E(k)]^2 = \lambda \quad (4)$$

Thus, the number of observed occurrences fluctuates about its mean λ with a standard deviation according equation (5)

$$\sigma_k = \sqrt{\lambda} \quad (5)$$

The variance as a function of k is the probability mass function. The Poisson distribution can be derived as a limiting case of the binomial distribution. The Poisson distribution can be applied to systems with a large number of possible events, each of which is rare. A classic example is the nuclear decay of atoms. The Poisson distribution is sometimes called a Poissonian, analogous to the term Gaussian for a Gauss or normal distribution.

Assumptions of Poisson distribution are:

- Observations are independent.
- Probability of occurrence in a short interval is proportional to the length of the interval.
- Probability of another occurrence in such a short interval is zero.

We verify that this Poisson distribution belongs to the exponential family as defined by Nelder and Wedderburn (1972). By taking logs of the Poisson distribution function, we find

$$\log f_i(y_i) = y_i \log(\mu_i) - \mu_i - \log(y_i!) \quad (6)$$

Looking at the coefficient of $y_i$we see immediately from (7) that the canonical parameter is

$$\theta_i = \log(\mu_i) \quad (7)$$

and therefore that the canonical link is the log. Solving for $\mu_i$we obtain the inverse link

$$\mu_i = e^{\theta_i} \quad (8)$$

and we see that we can write the second term in (14) the p.d.f. as

$$b(\theta_i) = e^{\theta_i} \tag{9}$$

The last remaining term in (14) is a function of y_i only, so we identify

$$c(y, \varphi_i) = \log(y_i!) \tag{10}$$

Finally, note that we can take $a_i(\varphi)$and $\varphi = 1$, just as it is in the binomial case. Let us verify the mean and variance.

Differentiating the cumulant function $b(\theta_i)$ we have

$$\mu_i = b(\theta_i) = e^{\theta_i} \tag{11}$$

And differentiating again regarding equation (14) we have

$$v_i = a_i(\varphi)b''(\theta_i) = e^{\theta_i} = \mu_i \tag{12}$$

Hence the mean is equal to the variance

The Exponential Family

GLMs may be used to model variables following distributions in the exponential family with probability density function

$$f(y; \theta, \varphi) = exp\left\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y; \varphi)\right\} \text{or}$$

$$\log f(y; \theta, \varphi) = exp\left\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y; \varphi)\right\} \tag{13}$$

Where φ is a dispersion parameter and $a(\varphi)$, $b(\theta)$ and $c(y; \varphi)$ are known functions in equations (14).
For distributions in the exponential families, the conditional variance of Y is a function of the mean, μ together with a dispersion parameter φ.

That is, $E(Y_i) = \mu_i = b'(\theta)$ and

$$var(Y) = \sigma_i^2 = b''(\theta)a(\varphi) \tag{14}$$

Where $b'(\theta)$ and $b''(\theta)$ are the first and second derivatives of $b(\theta)$. The dispersion parameter is usually fixed to one for some distributions.

Poisson Regression

The Poisson regression model assumes that the sample of n observations y_i are observations on independent Poisson variables $Y_i$with mean μ_i. Note that, if this model is correct, the equal variance assumption of classic linear regression is violated, since the Y_i have means equal to their variances.

So we fit the generalized linear model,

$$\log(\mu_i) = x_i\beta \tag{15}$$

We say that the Poisson regression model is a generalized linear model with Poisson error and a log link so that

$$\mu_i = \exp(x_i\beta) \tag{16}$$

This implies that one unit increases in an $x_i$are associated with a multiplication of $\mu_i$by $exp(\beta_j)$.

Model specification

The primary equation of the model is

$$P(Y_i = y_i) = \frac{e^{-\mu}\mu^{y_i}}{y!}, y_i = 0,1,2, \dots \quad (17)$$

The most common formulation of this model is the log-linear specification:

$$\log(\mu_i) = x_i\beta \quad (18)$$

From (3.20) the expected number of events per period is given by

$$E(y_i/x_i) = \mu_i = e^{x_i\beta} \quad (19)$$

Thus:

$$dE(y_i/x_i) = \beta e^{x_i\beta} = \beta_i\mu_i \quad (20)$$

The major assumption of Poisson model is

$$E(Y_i/x_i) = \mu_i = e^{x_i\beta} = var(Y_i/x_i) \quad (21)$$

This assumption would be tested later on. If $var(Y_i/x_i) > E(Y_i/x_i)$ then there is over-dispersion. If $var(Y_i/x_i) < E(Y_i/x_i)$, then under-dispersion has occurred.

Estimation

Estimation involves estimating the regression parameters specifically using the maximum likelihood estimation.

Maximum Likelihood Estimation

The likelihood function for n independent Poisson observations is a product of probabilities given by

$$\Pr(y_i) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}, i = o, 1,2, \dots \quad (22)$$

Taking logs and ignoring a constant involving $log(y_i)!$ we find that the log-likelihood function is

$$\log L(\beta) = \sum_{i=1}^{n}[-\lambda_i + y_i x_i\beta - log(y_i)!] \quad (23)$$
$$= \sum_{i=1}^{n}\left[-e^{x_i\beta} + y_i x_i\beta - log(y_i)\right] \quad (24)$$

where$y_i = \mu_i = e^{x_i\beta}$ (25)

The parameters of this equation can be estimated using maximum likelihood method

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n}\left(y_i - e^{x_i\beta}\right)x_i = 0 \quad (26)$$

and

$$\frac{\partial^2 L}{\partial\beta\partial\beta} = \sum_{i=1}^{n}\left[e^{x_i\beta}x_i x_i\right] \quad (27)$$

this is the Hessian of the function and with typical element

$$\frac{\partial^2}{\partial\beta_j\partial\beta_i} = \sum_{i=1}^{n}\left[e^{x_i\beta}x_{ij}x_{il}\right]; j.l = 1,2, \dots, p \quad (28)$$

As

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_i} = -\sum_{i=1}^{n}\left[e^{x_i\beta} x_{ij} x_{il}\right] \qquad (29)$$

does not involve the y data

$$k_{jl} = E\left(\frac{\partial^2 L}{\partial \beta_j \partial \beta_i}\right) = -\sum_{i=1}^{n}\left[e^{x_i\beta} x_{ij} x_{il}\right]; j.l = 1,2,\dots,p \qquad (30)$$

and the information matrix is

$$K = \sum_{i=1}^{n}\left[e^{x_i\beta} x_i x_i\right] \qquad (31)$$

There is no closed form solution to, $\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n}\left(y_i - e^{x_i\beta}\right)x_i = 0$ so the MLE for β must be obtained numerically.

However, as the Hessian is negative definite for all $x$and β, the $MLE(\hat{\beta})$ is unique, if it exists.

From $\frac{\partial^2 L}{\partial \beta_j \partial \beta_i} = -\sum_{i=1}^{n}\left[e^{x_i\beta} x_{ij} x_{il}\right]$ and:

$$k_{jl} = E\left(\frac{\partial^2 L}{\partial \beta_j \partial \beta_i}\right) = -\sum_{i=1}^{n}\left[e^{x_i\beta} x_{ij} x_{il}\right] \qquad (32)$$

$$k_{jlr} = E\left(\frac{\partial^2 L}{\partial \beta_j \partial \beta_i \partial \beta_r}\right) = -\sum_{i=1}^{n}\left[e^{x_i\beta} x_{ij} x_{il} x_{ir}\right] \qquad (33)$$

and

$$k_{jl}^{(r)} = \left(\frac{\partial k_{jl}}{\partial \beta_r}\right) - \sum_{i=1}^{n}\left[e^{x_i\beta} x_{ij} x_{il} x_{ir}\right], j, l, r = 1,2,\dots,p \qquad (34)$$

To make matters more transparent, consider the case of a single covariate and an intercept. Then is a scalar observation and

$$L = \sum_{i=1}^{n}[-\lambda_i + y_i x_i(\beta_1 + \beta_2 x_i) - log(y_i)!] \qquad (35)$$

where

$$\lambda_i = \exp(\beta_1 + \beta_2 x_i)\text{for} i = 1,2,\dots,n \qquad (36)$$

The first order conditions, $\frac{\partial L}{\partial \beta} = 0$ yield a system of K equations (one for each β) of the form

$$\sum_{i=1}^{n}\left(y_i - e^{x_i\beta}\right)x_i = 0 \qquad (37)$$

Where $y_i = e^{x_i\beta}$is the fitted value of y_i. The predicted/fitted value has as usual been taken as the estimated value of $E(\frac{y_i}{x_i})$. This first order condition tells us that the vector of residual is r orthogonal to the vectors of explicative variables.

3. DATA ANALYSIS AND RESULTS

This section introduces the analysis of the various models and discussion of findings. Preliminary analysis, summary of results and snapshot of the data will also be presented and discussed. Poisson regression model shall be used in the modeling.

Source of Data

The data to be used will be obtained from the State Hospital, Ibadan. The data will be on HIV/AIDS infection rate and as where the following characteristics of the AIDS victims like gender, marital status, occupation, and mode of transmission.

Data Presentation

A routine time data was obtained from State Hospital, Ibadan from 2014 to 2018. The data obtained is presented below

Table 1: Incidence of HIV in State Hospital, Ibadan, 2014

Month	Gender		Marital Status		Mode of Transmission, 1=Sex; 2=Others
	Male	Female	Married	Single	
JAN	12	39	50	1	1
FEB	18	41	59	0	1
MAR	17	42	59	0	1
APRIL	22	41	60	3	1
MAY	18	32	50	0	1
JUN	15	40	53	2	1
JULY	18	43	60	1	1
AUG	16	47	60	3	1
SEP	13	31	42	2	1
OCT	13	33	46	0	1
NOV	11	34	43	2	1
DEC	12	24	36	0	1

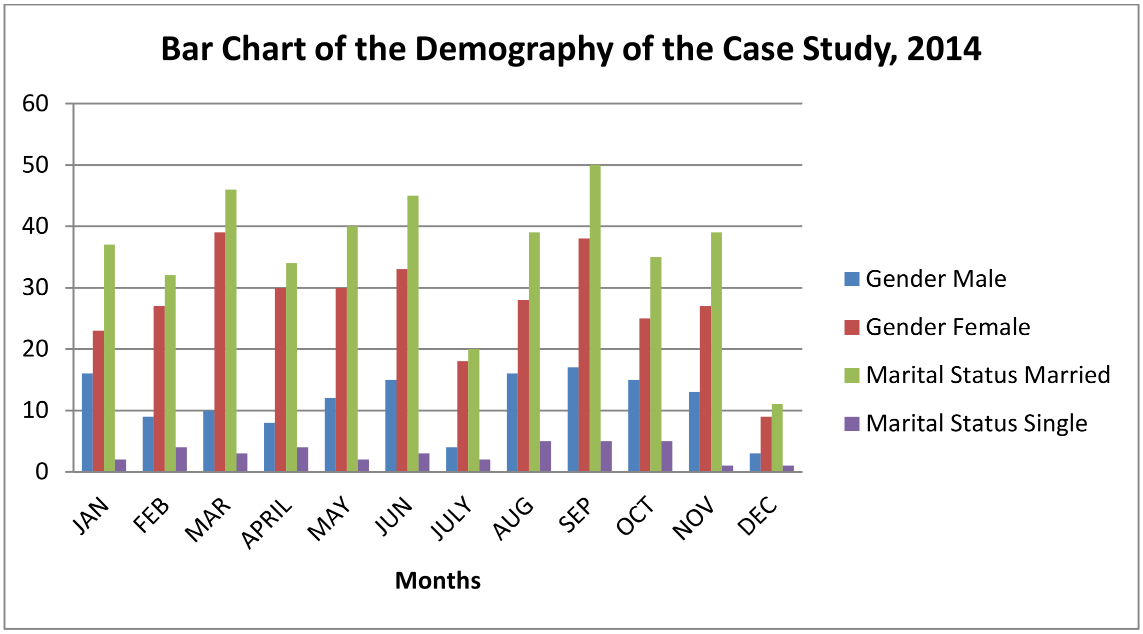


Table 2: Incidence of HIV in State Hospital, Ibadan, 2015

Month	Gender		Marital Status		Mode of Transmission 1=Sex; 2=Others
	Male	Female	Married	Single	
JAN	12	37	48	1	1
FEB	10	45	44	1	1
MAR	20	41	59	2	1
APRIL	21	42	63	0	1
MAY	19	50	69	0	1
JUN	15	38	52	1	1
JULY	15	31	45	1	2
AUG	26	45	70	1	1
SEP	17	50	67	1	1
OCT	13	27	40	0	1
NOV	10	38	48	0	1
DEC	9	34	40	3	1

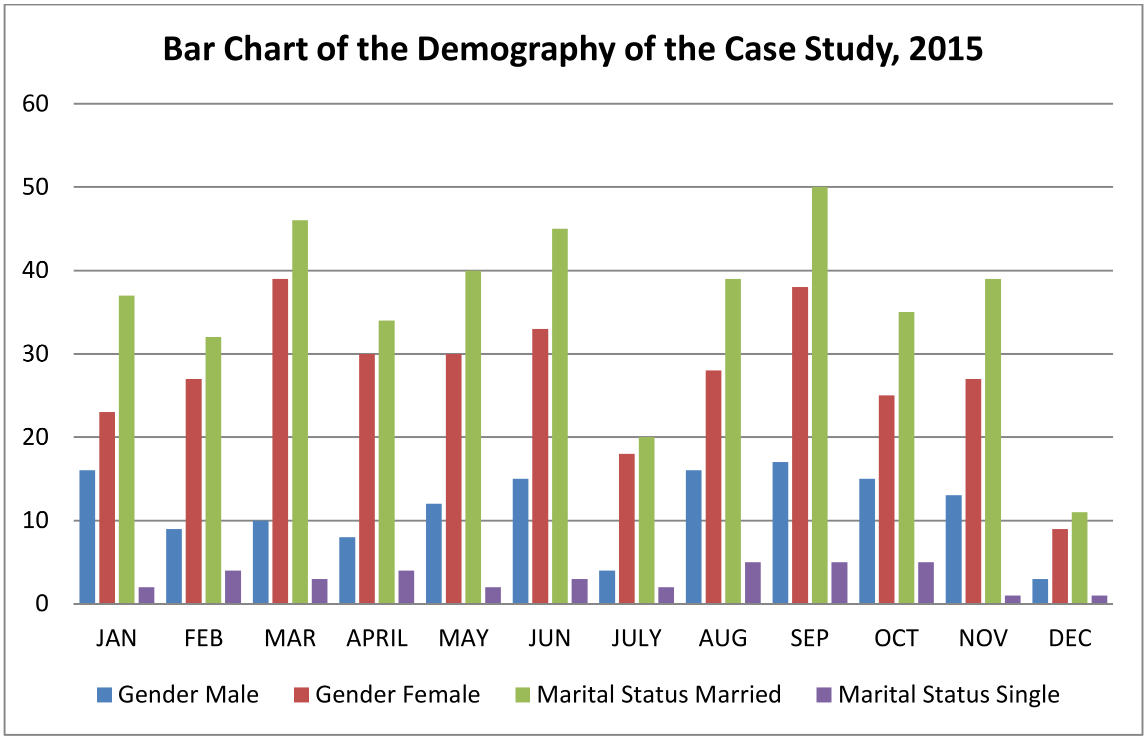


Table 3: Incidence of HIV in State Hospital, Ibadan, 2016

Month	Gender		Marital Status		Mode of Transmission 1=Sex; 2=Others
	Male	Female	Married	Single	
JAN	14	30	44	0	1
FEB	16	55	71	0	1
MAR	9	34	43	0	1
APRIL	15	42	57	0	1
MAY	22	36	58	0	1
JUN	10	39	49	0	1
JULY	14	48	62	0	1
AUG	11	46	57	0	1
SEP	26	46	72	0	2
OCT	16	44	60	0	2
NOV	11	46	57	0	1
DEC	11	35	46	0	1

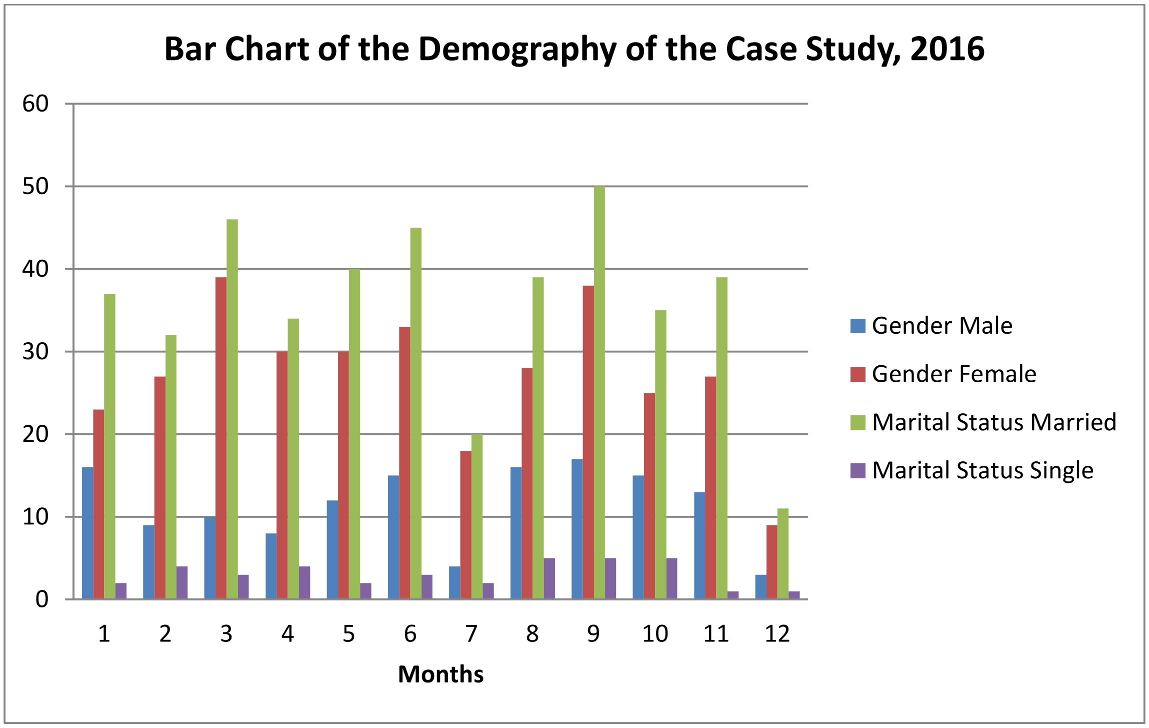


Table 4: Incidence of HIV in State Hospital, Ibadan, 2017

Month	Gender		Marital Status		Mode of Transmission 1=Sex; 2=Others
	Male	Female	Married	Single	
JAN	17	50	61	6	1
FEB	15	38	50	3	1
MAR	8	35	41	2	1
APRIL	12	30	40	2	1
MAY	14	39	53	0	1
JUN	14	28	42	0	1
JULY	18	46	59	5	1
AUG	19	44	62	1	2
SEP	15	23	37	1	1
OCT	16	26	40	2	1
NOV	13	29	38	4	1
DEC	16	31	44	3	1

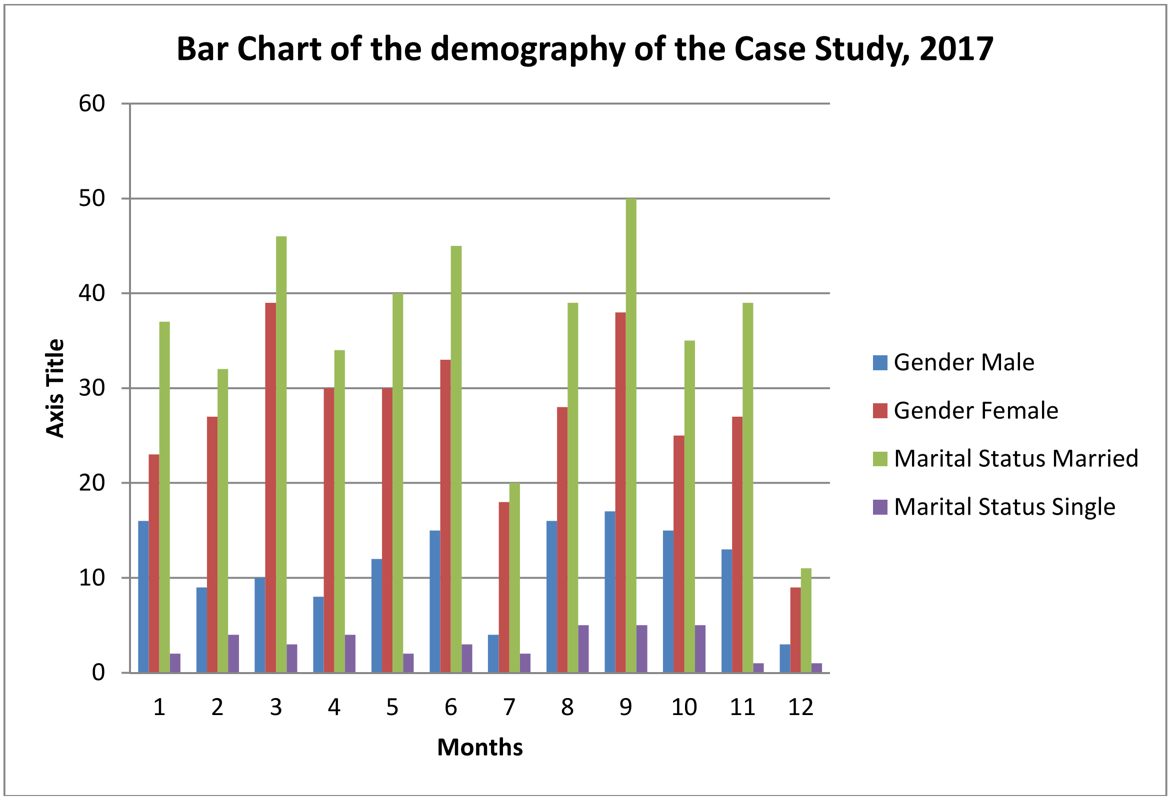
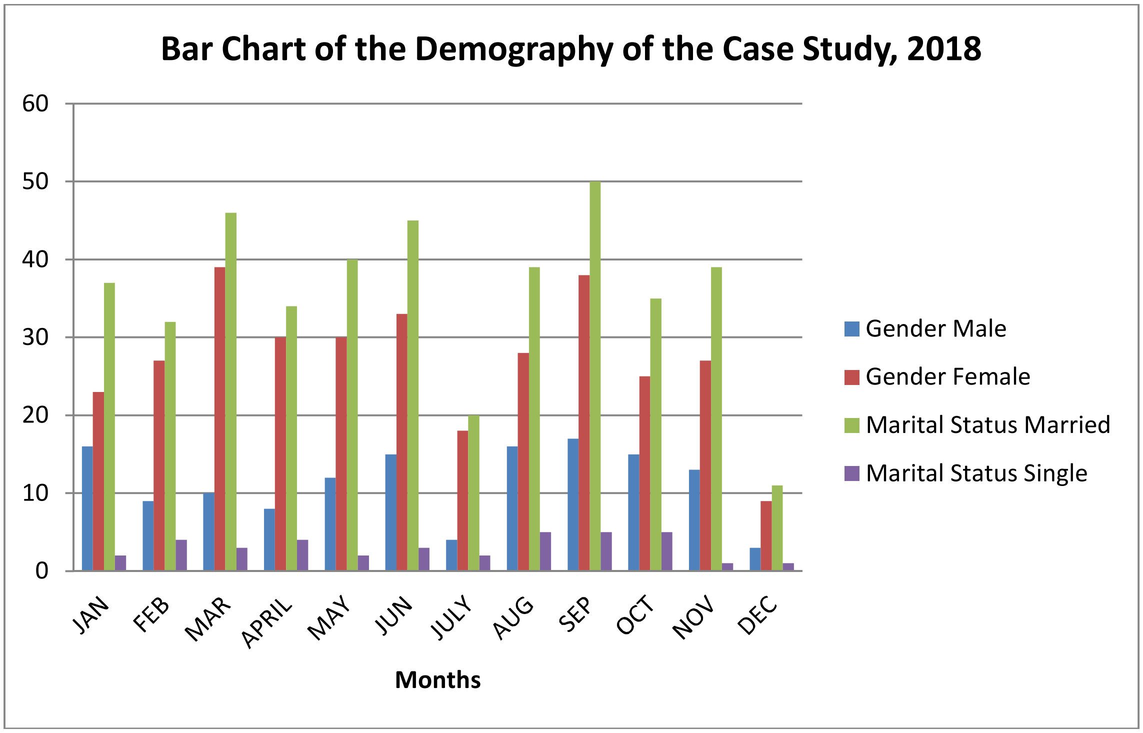


Table 5: Incidence of HIV in State Hospital, Ibadan, 2018

Month	Gender		Marital Status		Mode of Transmission 1=Sex; 2=Others
	Male	Female	Married	Single	
JAN	16	23	37	2	1
FEB	9	27	32	4	1
MAR	10	39	46	3	1
APRIL	8	30	34	4	1
MAY	12	30	40	2	1
JUN	15	33	45	3	1
JULY	4	18	20	2	1
AUG	16	28	39	5	1
SEP	17	38	50	5	1
OCT	15	25	35	5	1
NOV	13	27	39	1	1
DEC	3	9	11	1	1



Modeling and Criteria for assessing Model Goodness of Fit

The model goodness of fit assessment criteria results as outputted from the R software using the GLM procedure for Poisson regression models are shown in Table 6 and its coefficients estimate are also depicted in Table 7 with HIV cases being the response variable given the independent variables gender, marital status and mode of transmission. Again let α_1, be the respective intercept estimates where i=1, 2 represent the various intercept for the four models and β_i, β_j and β_l denote the estimates of the independent variables for i= 1,2,3 and 4; j=0 and 1 and l=1,2,3,4 representing gender, quarters, marital status and mode of transmission.

Modeling the Occurrence of HIV Incidence

Table 6: Poisson Regression Models with their AIC's

Models	AIC
$\boldsymbol{In(meancase) = \alpha_1 + \beta_k gender_i}$	16.987
$\boldsymbol{In(meancase) = \alpha_1 + \beta_k quarters_i}$	20.654
$\boldsymbol{In(meancase) = \alpha_1 + \beta_k maritalstatus_j, 1}$	23.974
$\boldsymbol{In(meancase) = \alpha_1 + \beta_k modeoftrasmission_l}$	17.987
$\boldsymbol{In(meancase) = \alpha_1 + \beta_k gender_i + \beta_m^{**} maritalstatus + \beta_n^{***} ModeofTransmission,}$	16.638

From Table 6, model 4 was chosen to assess the goodness of fit test because it satisfied all the assumptions with an AIC value of 16.638, a deviance of 14.505 on 36 degrees of freedom following the chi-square distribution $\chi^2_{(1)}$.The corresponding p-value associated with this model is < 0.00 and this indicates over-dispersion. Table 7 shows the parameter estimates of the selected Poisson regression model for model 4.

Table 7: Parameter Estimates of the Selected Poisson Model

Coefficients	Estimates	Standard errors	z values	pr(>|z|)
intercept	5.246723	0.015673	337.962	<0.001
quarters1	-0.12346	0.012778	-23.111	<0.001
quarters2	-0.148015	0.012591	-11.755	<0.001
quarters3	-0.083806	0.02379	-6.77	<0.001
quarters4	-0.239595	0.012912	-18.557	<0.001
female	0.047856	0.009074	5.224	<0.001
male	-0.644854	0.022304	-28.912	<0.001
Married	-0.113363	0.019051	-5.95	<0.001
Single	-0.0607	0.018792	-3.23	0.001237
Mode	-0.039282	0.018698	-2.102	0.03559

A dispersion parameter of 47.40196 shows that the data is over-dispersed i.e. a situation where the variance of the response variable exceeds the mean. In the nut shell, Poisson regression model can fit the data. Table 4.7 depicts the parameter estimates after validating the Poisson regression. The model obtained is given as:

$$
\begin{aligned}
In(mean\ case) = {} & 5.246723 + 0.047856 female - 0.644854 male - 0.113363 married \\
& - 0.0607 single - 0.12346 quarter1 - 0.148015 - 0.083806 - 0.239595 \\
& - 0.039282 mode\ of\ Transmission
\end{aligned}
$$

4. CONCLUSION

The objective of this research was to model the prevalence of HIV/AIDS cases given the age, gender and time (quarters); to model the incidence of severe HIV/AIDS cases given the age, gender and time in years and lastly to validate the two models using negative binomial regression model. Data from State Hospital, Ibadan were utilized in this study. Severe HIV/AIDS cases confirmed by the laboratory, simple HIV/AIDS (laboratory confirmed) and simple HIV/AIDS (non-laboratory confirmed) were used in the analysis and modeling. Poisson regression model was used and well known statistical goodness of fit model assessment criteria were used in selecting which model will fit the HIV/AIDS cases better. Based on the results, the Poisson regression model was found to fit the data better. In modeling the occurrence of HIV/AIDS cases, the analysis produced a reasonable AIC values, (16.638) deviance (14.505); p-value <0.00 for the Poisson model and a dispersion parameter of 47.40196 showing an extra-Poisson variation The occurrence of HIV/AIDS cases in quarter4 (October- December) was found to be e2734. 0 (0.7607) accounting for 76% of all laboratory confirmed cases recorded between 2014 to 2018.

50% laboratory confirmed cases were children below 5 years with more cases recorded for 20-34 age groups accounting for (48%) and those found in 20-34 age groups recording the highest number of cases accounting for 62% of laboratory confirmed cases. Again the Poisson regression model fitted the data. The variable time (in years) had a coefficient of 0.3470 which is statistically significant at 5% α-level.

This means that for each one-unit increase in time (years), the expected log count of the incidence of severe HIV/AIDS cases will increase by 0.35262 (i.e.) e35262. 0(1.4227). It also follows that as time goes on there will be an increase in severe HIV/AIDS cases.

Poisson regression also showed the independence of HIV/AIDS with respect to gender.
With reference to the findings of the research, it can be concluded that:

1. The occurrence of HIV/AIDS cases and the incidence of severe HIV/AIDS cases were independent of gender.
2. The occurrence of HIV/AIDS cases was found to be very high in the last quarter (October-December) between 2014 to 2018.
3. Between 2010 to 2014, the incidence of severe HIV/AIDS cases increased.
4.

On the basis of the findings of the research, the following recommendations were made:

- Since more cases were recorded in the last quarter (October-December) of the years considered apparently due to some seasonal changes, it is imperative that programmes and campaigns meant to reduce the menace of HIV/AIDS should be carried out before, during and after the seasonal changes.
- The education on the awareness of the causes as well as the preventive measures of HIV/AIDS should be carried out immensely in all nooks and cranny of state.
- The State AIDS agency should as part of its corporate social responsibility in the municipality must intensify its campaign in order to bring the disease under control.
- It is therefore suggested that more studies and research be carried out in highly utilized hospital in other to check the incidence of HIV/AIDS so that appropriate measures and strategies could be adopted to curtail its spread.

REFERENCES

Cameron, A.C. andTrivedi, P. K. (2008), Regression Analysis of Count Data, econometric Society Monographs, No. 30, Cambridge University Press.

Chunag.K.J.,Chan.C.C., Lee.C.T., and Tang.C.S. (2007). The effect of urban air pollution on inflammation, oxidative stress, coagulation and autonomic dysfunction in young adults. *AM Journal Respiration Care Medicine*, 176: 370–376.

McCullagh, P. &Nelder, J. (2009), Generalized Linear Models, Second Edition, Monographs on Statistics and Applied Probability Series, London: Chapman & Hall.

Morgan.G.,Corbett.S., Wlodarczyk.J., and Lewis.P. (2008). Air pollution and daily mortality in Sydney, Australia, 1989 through 1993.*American Journal of Public Health*, 88(5):759–764.

Parodi S and Bottarelli E. (2006) Control for confounding in case-control studies.*AnnalidellaFacoltà di MedicinaVeterinaria di Parma*, XXV, 14-46.

WHO (2004), Expect Committee on Malaria-Twentieth Report. World Health Organisation, Geneva.