Generalized Linear Model (GLM) Approach Using Poisson Probability Distribution on the Epidamic of HIV/AIDS in Kebbi State

James T.O[1], Manga, U. S.[2]

Abstract: HIV/AIDS epidemic is a real menaces world over and Kebbi state in particular. In this study we used HIV/AIDS surveillance data from National HIV/AIDS and Reproductive Health Survey (NARHS), Federal Ministry of Health, to model the situation in the state within the period of six years. UNAIDS estimation and EPP were used. The deviance analysis and AIC were employed to measure the goodness of fit of the models. There were 2570 diagnosed patients (804 male and 1706 female).Our results shows that the models fit very well because of the smaller values of the AICs (57.978), (the smaller the value of the AIC the better the goodness of fit of the model). The highest prevalence among age-group (23.8% and 21.8% for age-group 31-45 and 46-60 respectively), the epidemic prevalence peaked is 2002 (25.6%) and 2000 (23.6%), the incidence rate is 23.7% for GHY and 22.6% GHZ, indicating the pandemic in southern part of the state. The commonness of the pandemic is the female 66.5% as compared to male 33.5%.

Keywords: AICs, Epidemiology, GLM, HIV/AIDS, Poisson Probability Distribution.

1. INTRODUCTION

Human Immunodeficiency Virus (HIV) is the causative organism of AIDS which has become one of the greatest public health challenges faced by mankind. AIDS was first identified in 1981 in Los Angeles, USA. Two types of HIV exist presently-HIV-1 and HIV-2 (Alizon et al., 2010; Adoga et al., 2010). From late the 1970s to early 1980s in the United States and Europe, the emergence of immunologic dysfunction among patients of unknown etiology attracted popular attention [David et al, 2001]. Following the unusual occurrence of Pneumocystis carinii pneumonia (PCP) in five homosexual men from Los Angeles was reported to the Centers for Disease Control and Prevention (CDC) in 1981 [Anonymouse, 1989], several similar reports, describing male homosexuals and intravenous drug users with impaired immune systems and T lymphocytes were sent to CDC [Gottlieb et al, 1981]. In 1983, scientists at the Pasteur Institute discovered a virus from the lymph nodes of an asymptomatic individual; the scientists presented their discovery with Gallo and colleagues at the National Institutes of Health [Barr-Sinoussi et al, 1983]. Subsequently Gallo reported the isolation of retroviruses from AIDS patients, which they named HTLV-III. In 1994 Levy et al reported a similar retrovirus isolated from both AIDS patients and healthy individuals from the various risk groups [Levy et al, 1994] which they named the AIDS-associated retrovirus (ARV). The new retrovirus, associated with AIDS in the United States, Europe, and central Africa and exhibiting typical morphologic andgenetic characteristics of the Lentivirus genus, was named human immune deficiency virus (HIV) [Coffin et al, 1986], and subsequently HIV-1. In 1986, a related, but

ISBN - 978-93-81583-57-9

immunologically distinct and less pathogenic human retrovirus (now called HIV-2), was recovered from individuals residing in several Western African countries [Clavel et al, 1986].

1.1 Epidemiology: The HIV pandemic is one of the most notorious infectious disease epidemics in Human history. Its morbidity and mortality rates are staggering. In 2001, there were 36 Million HIV-infected individuals worldwide [FMOH, 2005]. In 2005, it was estimated that 38.6 (range35~46) million people worldwide were living with HIV. More than 21.8 million deaths have been due to HIV infection since the beginning of the epidemic [FMOH, 2005]. In 2005 alone, 3.4~4.2 million people died of AIDS and nearly 4.1 (range 3.4~6.2) million individuals worldwide acquired HIV infection [FMOH, 2005]. There were an estimated 925,000-1,025,000 persons living with HIV/AIDS at the end of 2003 in the US, and approximately 40,000 new HIV infections occur each year [Glynn, 2005]. HIV is transmitted by:3 ways

- Mucosa contact (oral, rectal, or vaginal) during sex;
- Transfusion of HIV contaminated blood products, use of contaminated equipment; and
- Maternal-fetal circulation or by breast feeding.

Sexual transmission accounts for more than 90% of HIV infections worldwide [Glynn, 2005]. Transmission of HIV is dependent on behavioral and biologic factors. The probability of male-to-female HIV transmission during vaginal sex is approximately 0.1% to 0.2% per contact; receptive anal intercourse is associated with a considerably higher risk (0.82/per contact) of HIV transmission [Vittinghoff et al, 1999]. Usually, HIV infection includes a long period (approximately 10 years) of clinical latency between the time of primary infection and the development of symptoms indicative of advanced immunodeficiency. Kebbi State with a populations of 3.2 million (2006 National Census) has a prevalence rate of 4%. While the country's prevalence rate has reduced from 5% in 2003 to 4.4% in 2005, the State prevalence increased from 2.5% to 4% during the period under review. It is higher in the rural areas than the urban areas. However, out of the 75,000 people that were infected, only 5,000 of them had submitted themselves for treatment, (FMOH, 2005)

2. METHODS

The generalized linear modelling structure comprises a linear combination of predictor variables related via a link function to the mean of the response distribution selected from the exponential family of distributions (Paul and Rayner 2011). We examine the application of Generalized Linear Models through the poisson probability distributions model to investigate the epidemiology of HIV/AIDS in Kebbi State base on age-group and health centers within six years. The selected health centers are General Hospital, Zuru (GHZ), Sir Yahaya Hospital, Birnin Kebbi (SY), General Hospital, Argungu (GHAR), General Hospital, Aliero (GHAL), General Hospital, Yauri (GHY) and Federal Medical Centre, Birnin-

Kebbi (FMC). Using the data collected by National HIV/AIDS and Productive Health Survey (NARHS) Nigeria, to look into the situation of HIV/AIDS in the state. We shall implement R 2.10.1 software package for the data analysis of the research. The state has four major tribes which include: Hausa, Dakarkari, Fulani and Gungawa. Islam is the dominant religion of the people; other tribes are Bussawa Dukkawa, Kambari and Kamuku ethnic communities. The major occupations of the people are Civil-Servants, Famers, Cattle-Rearers and Traders. The 21 local Governments are Aleiro, Arewa, Augie, Argungu, Bagudo, Birnin-kebbi, Bunza, Dandi, Danko-wasagu, Fakai, Gwandu, Jega, Kalgo, Koko-Besse, Maiyama, Ngaski, Sakaba, Shanga, Suru, Yauri and Zuru, (FMOH, 2005). In the GLM framework, it is customary to use a quantity known as deviance to formally asses Models adequacy and to compare models. Deviance statistics are identical to those obtained Using Likelihood ratio test statistics.

2.1 The Model Design: If $y_1, \ldots, y_n$ denote n independent observations on a response of HIV/AIDS in Kebbi State. We treat y_i as a realization of a random variable. In the generalized linear model we assume that Y_i has a normal Distribution with mean μ_i and variance σ^2

$$y_i \sim N(\mu_i, \sigma^2) \qquad 2.0$$

And we further assume that the expected value μ_i is a linear function of p predictors that Take values $x_i^I = (x_{i1}, \ldots, x_{ip})$ for the $i-th$ case, so that

$$\mu_i = x_i^I \beta, \qquad 2.1$$

Where β is a vector of unknown parameters

We will generalize this in two steps, dealing with the stochastic and systematic components of the model. We assume that the population from which the sample data are obtained can be stratified into H strata within each stratum h, n_h cluster or primary sampling unit (PSUs) are drown And within the h_t Stratum and its cluster nh_i ultimate sampling units (USUs) are drown sex, health-centers) a with Design weight s where j Denote the jth USU within the ith cluster, which in turn is

nested Within stratum h, Furthermore, we assume that the row of the matrix Y are represented by

$$n = \sum_{h=1}^{k} \sum_{i=1}^{n} n_{hi} \qquad 2.2$$

Observation of the p response variable Y and that the rows of the matrix X are n observations of the explanatory variable x. Thus

$$Y = E[Y] + \epsilon$$

where $E[Y] = XB$

$$\mu_i = E[Y] = g^{-1}(\Sigma_I X_{ij} B_I + \beta_I) \qquad (2.3)$$

Where
$Var(Y) = \phi V(U_i)/W_i$

and
Y_i Is the vector of responses of HIV/AIDS

$g(x)$ is the link function which relate the expected responses to the linear combination of observe factors.
X_{ij} Is a matrix (the "design matrix") produce from the factors of the sexes.

B_j Is the vector of model parameters which is to be estimated

β_I Is a vector of known effects of the HIV/AIDS

ϕ Is a parameter to scale the function

V(x) is the variance function.
W_I Is the prior weight that assigns credibility, or weight to each observation. Since the age-group data collected for HIV/AIDS is for the period of six years (2000-2005) in Kebbi State is a counts data, then the Poisson Probability Distribution Model is a contender thus the probabilities that a random variable Y that Describes counts taken on values in the range 0, 1, 2, 3, ... more precise. Has the likelihood tc be

$$L(y) = \prod_{i,j} \frac{e^{-u_{i,j}} u_{i,j}^{y_{i,j}}}{y_{i,j}!} \tag{2.4}$$

If we constrain $\sum_{I,J} \mu_{I,J} = N$ then this can be

$$L(y) = \frac{e^{-N} N^{\mu}}{\prod_{I,J} y_{i,j}!} \prod_{i,j} p_{i,j}^{y_{i,j}}$$

Then the likelihood for this equation has the form

$$\sum_{i,j} \hat{\mu}_{i,j} = \sum_{i,j} y_{i,j} \tag{2.5}$$

For a two-way contingency table we are commonly interesting whether the row and column factors are independent under the standard rule of probability theory the hypothesis is

$$p_{i,j} = p_{i.} p_{.j} \qquad where\ p_{i.} = p_{.j}$$

Are the row and column marginal probabilities respectively. Relating to the Poisson mean we have

$$\mu_{i,j} = N p_{i.} p_{.j}$$

Hence

$$\log(\mu_{i,j}) = \log(N) + \log p_{i.} + \log p_{.j}$$

$$= c + a_i + b_j$$

Thus the full model is

$$\Rightarrow \log(\mu_{i,j}) = c + a_i + b_j + (ab)_{i,j} \tag{2.6}$$

Since the data produced is by clustered Poisson process where each event contribute a random amount to the total. Then c, a_i b_j and $(ab)_{ij}$ are independent and identically distributed, and the deviance for the Poisson model is

$$\therefore D = \sum_{I} \{ y_i \log \frac{y_i}{\hat{\mu}} - (y_i - \hat{\mu}_i) \} \tag{2.7}$$

3. RESULTS AND DISCUSSION

The prevalence distribution of HIV/AIDS among age group in the six selected health centers in Kebbi State is for the period of six years. Is therefore a count data and required the Poisson distribution model, (Equation 2.6). Table I, shows the prevalence distribution of HIV/AIDS among the six selected health centers in Kebbi State.

Table I. The Prevalence Distribution of HIV/AIDS (Positive Cases) Among Age-Group in the six selected Health Centres in Kebbi State

Year/Age-Group	<1	1-15	16- 30	31 -45	46- 60	>60
2000	109	77	132	139	136	7
2001	86	47	73	71	75	5
2002	109	66	172	166	133	13
2003	71	55	81	79	81	11
2004	55	50	44	76	62	9
2005	60	27	57	81	48	7
Total	490	322	559	612	535	52
Prevalence(%)	2.0	19.1	12.5	23.8	21.8	20.8

For each category from table. I above, the first level was used as the reference level. For age-group, HIV/AIDS infected patients have the highest relative prevalence hazard rate (23.8% and 21.8% for 31-45 and 46-96, respectively). HIV/AIDS cases had the lowest relative prevalence hazard rate (2.0 and 12.5 for <1 and 16-30, respectively).

3.1 Fitting the HIV/AIDS Age-Group Data in the Poisson Probability Distribution Model: Using the population as a measure of size (MOS), a cumulative population of the age-group with probability proportional to the size for each stratum, the cumulative population is with respect to the year (equation 2.6). Hence the result follows.

Table II. Data Frame of HIV/AIDS Age-Group

SNo.	Treatment	Age group	Counts
1	1	1	490
2	1	2	322
3	2	3	559
4	2	1	612
5	3	2	535
6	3	3	52

Table II, display the data frame of HIV/AIDS by age-group which is used in fitting the Poisson probability distribution model (equation 2.6).

Table III. Analysis of Deviance Table for HIV/AIDS Age-Group

	Df	Deviance Resid.	Df	Resid. Dev
NULL			5	701.10
agegroup	2	142.73	3	558.37
treatment	2	240.51	1	317.86
agegroup:treatment	1	317.86	0	1.226e-13

Table III shows the analysis of deviance. Model: Poisson, link: log, response: counts, terms added sequentially (first to last). We can see that the deviance residuals are significant, because of their smaller values with respect to the degree of freedom (Df).

Table IV. Deviance Residual for HIV/AIDS Age-Group

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.19441	0.04518	137.119	< 2e-16 ***
agegroup2	-0.41985	0.07174	-5.853	4.84e-09 ***
agegroup3	-2.75088	0.16201	-16.980	< 2e-16 ***
treatment2	0.22233	0.06062	3.668	0.000245 ***
treatment3	0.50772	0.07053	7.198	6.09e-13 ***
agegroup2:treatment2	NA	NA	NA	NA
agegroup3:treatment2	2.66029	0.17225	15.445	< 2e-16 ***
agegroup2:treatment3	NA	NA	NA	NA
agegroup3:treatment3	NA	NA	NA	NA ---

Considering the specification on the GLM relationship: glm (formula = counts ~ age group * treatment, family = poisson()) which is used to get the deviance residuals (Table IV). We can see from the table that the Deviance Residuals: [1] 0 0 0 0 0 0,

Dispersion parameter for poisson family taken to be 1, the null deviance: 7.0110e+02 on 5 degrees of freedom, Residual deviance: 1.2257e-13 on 0 degrees of freedom and AIC: 57.978. The significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1, which shows that the probability greater than the absolute value of z (Pr(>|z|)) is significant at 0%. The intercept value, estimate, standard error and the z value are also shown in the table Coefficients (NA): (3 not defined because of singularities).

Table V. Distribution Function Analysis for HIV/AIDS Age-Group

95% CI Approximate		95% CI Approximate	
Percent	Percentile	Lower Limit	Upper limit
1	-17.198	-309.993	275.60
2	34.618	-233.558	302.79
3	67.494	-185.582	320.57
4	92.225	-149.813	334.26
5	112.342	-120.952	345.64
6	129.465	-96.572	355.50
7	144.478	-75.349	364.31
8	157.921	-56.480	372.32
9	170.146	-39.437	379.73
10	181.400	-23.855	386.65

Table V shows the distribution parameter estimates on 95% CI with Variable: count, Mean: 425, StDev.: 190.082, Goodness of Fit: Anderson-Darling (adjusted) = 2.081.

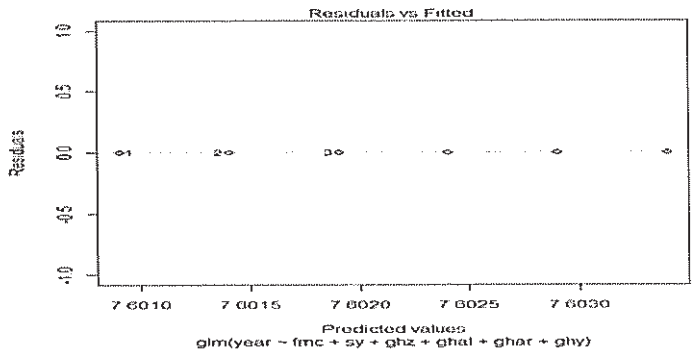

Fig 3.1 Residual vs fitted for HIV/AIDS on Health Centre

In fig 3.1, the residual vs. fitted was plotted and present the relative excess hazard rate for HIV/AIDS patients in all the health centers. The excess hazard rate of death for health centers over the period of study.

3.2 Conclusion: HIV/AIDS infection prevention is one of the most challenging tasks for public health workers. In order to improve the quality of life of HIV-infected patients and extend the interval from HIV/AIDS diagnosis to death, HIV/AIDS relative epidemiology analysis will provide a measure of the excess mortality experienced by patients diagnosed with HIV/AIDS infection, irrespective

of whether or not the excess mortality is directly or indirectly attributable to HIV infection.

Our results showed that the probability relative excess hazard of death was higher for age-group 31-45, with HIV/AIDS infection diagnosis as. Excess risk for death also was higher for persons with a diagnosis of HIV/AIDS at lower CD4+ T-cell count.

This is not surprising since similar methods and datasets were used to model the relative epidemiology of HIV/AIDS patient. The Model used grouped data (events were counted in approximate survival time). The resulting Goodness of fit test showed that our data fit well for the selected model because of the lower value of AIC's. Meanwhile, two term interactions were tested. It was found that there were significant interactions between the age-group. The main effects of follow-up year are significant in the model, health centers and diagnosis year.

Our results also showed that the estimated mean residual life for HIV/AIDS patients followed the same pattern as the survival probability: if the HIV/AIDS patient groups had a higher relative survival probability, then they had a longer mean residual life; if the patients had a lower survival probability, then they had a shorter mean residual life. The higher HIV/AIDS prevalence among women calls for focused and appropriate interventions to target this group.

4. REFERENCES

1. S. Alizon, V. Wyl, T. Stadler, D.R. Kouyos, S. Yerly, B. Hirschel, J. Boni, C. Shah, T. Klimkait, H. Furrer, A. Rauch, L.P. Vernazza, E. Bernasconi, M. Battegay, P. Burgisser, A. Telenti, F. H. Gunthard, S. Boenhoeffer. The Swiss HIV Cohort study (2010). *Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. PLOS pathogens,* volume 6 issue 9, e1001123.
2. H. Akaike. *"A new look at the statistical model identification"*. IEEE transaction on Automatic Control 19(6):716–723. doi:10.1109/TAC. 1974.1100705. MR0423716.
3. Anonymous. *Interpretation and use of the Western blot assay for serodiagnosis of human immunodeficiency virus type 1 infections.* MMWRMorb Mortal Wkly Rep 1989; 38:1–7.
4. F. Barr-Sinoussi, J.C. Chermann, and F. Rey. *Isolation of a T-lymphotrop retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science,1983; 220:868–871.
5. F. Clavel, D. Gutard, and F. Brun-Vzinet. *Isolation of a new human retrovirus from West African patients with AIDS.* Science, 1986; 233:343–346.
6. J. Coffin, A. Haase, and J. A. Levy. *Human immunodeficiency viruses* [letter].Science; 1986; 232:697.
7. M. K. David, M. Pever, and M. A. Martin. *Fields Virology, 4th edition,* Lippincott – Raven, Philadelphia , USA. Anonymous. Pneumocystis pneumonia: Los Angeles. MMWR Morb Mortal Wkly Re;30:250–252.doi:10.1016/j.spl.2006. 09.018 *among Married Men in Ile-Ife, Nigeria. Aging Male. 2001; 6 (2);7985.*
8. Federal Ministry of Health and World Health Organisation. (2005). *Plan to Scale-Up Antiretroviral Treatment for HIV or AIDS in Nigeria 2005-2009. Abuja, Federal Ministry of Health and World Health Organisation.*
9. M.K. Glynn, and P. Rhodes. *Estimated HIV prevalence in the United States at the end of 2003 [Abstract T1-B1101].* Proceedings of the 2005 National HIV prevention Conference; 2005 Jun 12-15; Atlanta, GA.

10. D. N. Levy, Y. Refaeli, and R. R. MacGregor. *Serum Vpr regulates productive Infection and latency of human immunodeficiency virus type 1*. Proc Natl Acad Sci U S A, 1994; 91:10873–10877.

11. P. McCullagh, J. A. Nelder. *Generalized Linear Models (2nd ed)*. Chapman & Hall: London, 1989.

12. P. Rippon, and J. Rayner. "Assessing Poisson and Logistic Regression Models Using Smooth Tests" (2011). Proceedings of the *Applied Statistics Education and Research Collaboration fourth annual ASEARC Conference;* 2011 February 17–18; Parramatta, Australia. Available online http://ro.uow.edu.au/asearc/21

13. E. Vittinghoff, J. Douglas, and F. Judson. *Per-contact risk of human immunodeficiency virus transmission between male sexual partners*. Am J Epidemiol; 1999; 150:306–311.

****

---

[1]*James, T.O.*
Department of Mathematics
Kebbi State University of Science and Technology, Aliero. Nigeria
Corresponding author: tolujam@yahoo.com
[2]*Manga, U. S.*
Department of Mathematics
Kebbi State University of Science and Technology, Aliero. Nigeria
usmathmanga@yahoo.com