New Poisson Regression Model as a Tool for Measuring Incidence of Maternal Death: Case Study of Ondo State Specialist Hospital, Akure, Nigeria

Mojeed, O. and Aladeniyi, O. B*
Federal University of Technology, Akure, Nigeria
*obaladeniyi@futa.edu.ng

ABSTRACT

In this work, we propose a new Poisson regression model for count data by incorporating the parameter (t) in the normal Poisson probability distribution function. The mean, variance, maximum likelihood estimation and Fisher's information for the proposed model were also formulated. Poisson regression model do not just analyze the data, but also tested for disparity in the mean and variance of count data. Verification of the proposed model was carried out on the data obtained from the Obstetrics & Gynecology directorate of Ondo State Specialists Hospital, Akure, Nigeria for the period 2005-2014 using the R package. The dispersion parameter (DP) is 1.0028, showing a clear absence of over dispersion in the data, also there exist a year to year variation in the occurrence of incidence of maternal death. Hence, the new Poisson model is appropriate for the data used.

Keywords: Mean, Variance, distribution function, dispersion

1.0 INTRODUCTION

The International Code of Diseases (ICD-10) defines maternal mortality as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes (WHO *et al*, 2007). At the close of the last century, sub-Saharan Africa still experienced high maternal mortality rates, with the goal of the provision of conditions conducive to safe motherhood eluding many governments (WHO *et al*, 2015). For many years, in spite of the efforts exerted in this regard, the evidence shows that the rate of maternal death was on the rise.

Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a given fixed interval of time or space if these events occur with a known average rate and independently of the time since the last event. Poisson distribution can be used to explain the behavior of the discrete random variables where the probability of occurrence of the event is very small and the total number of possible cases is sufficiently large. Poisson regression is similar to the regular multiple regressions except that the response variable (Y) is an observed count that follows the Poisson distribution. The aim of regression analysis in this instance is to model the dependent variable Y as the estimate of outcome using some or all of the explanatory variables. When the response variable is normally distributed we found that its mean could be linked to a set of explanatory variables using a linear function. A typical Poisson regression model expresses the log outcome rate as a linear function of a set of predictors.

Several renowned researchers in the field of Statistics had worked on incidence of maternal mortality using several statistical tools such as descriptive statistics and time series. While time series analysis shows the trend of occurrences and forecasts on the future occurrence base on the past data, Poisson regression model tends to analyze the data, test for disparity in the mean and variance of the data. Bucher *et al* (1976) examines racial differences in the incidence of ABO hemolytic disease by examining records for infants born at the North Carolina Memorial Hospital. Fisher et al (1922) considered the accuracy of the plating method of estimating the density of bacterial populations. Aitkin *el al*. (1990) and Renshaw (1994) each respectively fitted the Poisson model to two different sets of U.K. motor claim. The Poisson quasi likelihood model had been suggested to accommodate over-dispersion in claim count and this was applied to the count data of U.K. own damage motor claims by Brockman and Wright to take into account the possibility of within-cell heterogeneity (Noriszura & Abdul, 2007). Researchers like Smart & Abena (2013), Yunos *et al* (2016) and Zamani *et al* (2014, 2016) had worked on various Poisson models with applications to but not limited to count, survey and health care data.

This research work therefore formulating a new and an appropriate Poisson regression model to fit count data on maternal mortality by incorporating the variable t.

2.0 INTERMEDIATE SESSIONS

2.1 MODEL METHODOLOGY

Since maternal deaths are considered a count data, Poisson regression model was specified with the years as covariate, maternal death cases as the response variable and the time (in years) as the predictor variable. Class of generalized linear models (Nelder and Wedderburn, 1972) provided a unified framework to study various regression models such as the Poisson regression model. The Poisson model assumes that the variance of the count data is equal to the mean (Agrestin, 2007).

The Poisson distribution models the probability of y events (death) with the formula.

$$P_r(Y = y|\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \quad (y = 0,1, 2,\ldots)$$

The probability of y events is then

$$P_r(Y = y|\mu, t) = \frac{e^{-\mu t}(\mu t)^{y}}{y!} \quad (y = 0,1,2,\ldots)$$

2.2 NEW POISSON REGRESSION MODEL

From the fundamental of the linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k.$$

Logarithm of the response variable is linked to a linear function of explanatory variables $\log_e(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k and$ Y = $(e^{\beta 1x1})(e^{\beta 2x2}) \ldots (e^{\beta kxk})$.

This implies that, Poisson regression model expresses the log outcome rate as a linear function of a set of predictors

In Poisson regression, we suppose that the Poisson incidence rate μ is determined by k regressor variables (the X's).The expression relating in these quantities is

$$\mu = t\, exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k)$$

Note that often, X_1= 1 and β_1 is called the intercept. The regression coefficients $\beta_1, \beta_2, \ldots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are labeled $b_1, b_2, \ldots, b_k$.

Using this notation and incorporating the parameter t (time), the fundamental Poisson regression model for an observation is written as

$$P_r\ (Y_i = y_i|\mu_i t_i) = \frac{e^{-\mu_i t_i}(\mu_i t_i)^{y_i}}{y_i!}$$

Where: i=1,2,3….,10 (number of years of occurrence)

$$\mu = t_i \mu('\beta)$$

$$= t_i exp(\beta_{1i} X_{1i} + \beta_2 X_{2i} + \beta_{3i} X_{3i} + \cdots + \beta_{ki} X_{ki})$$

That is, for a given set of values of the regressor variables, the outcome follows Poisson distribution.

Hence the Poisson regression model used in this study is given by

$$Y \sim Possion[t_i\mu, (X_i, \beta)] i = 1,2,\ldots,n$$

Where t is the time length in which events occur, which is known.

Hence the probability distribution function of the Poisson regression model is given by

$$P_r(Y_i = y_i) = \frac{(t_i\mu(X_i,\beta))^{y_i}}{y_{i!}} e^{-t_i\mu(X_i,\beta)} \quad (1.0)$$

Where

$$\mu(X_i, \beta) = e^{Xi^t\beta} \geq 0 \quad (1.1)$$

$$E[Y] = t_i\mu(X_i, \beta) = var[Y] \quad (1.2)$$

2.3 FORMULATION OF MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood estimation is formulated as follows:

From equation (1.0)

$$L(y,\beta) = \prod_{i=1}^{n} P_r(Y_i = y_i) = e^{-\sum_{i=1}^{n} t_i\mu(X_i,\beta)} \prod_{i=1}^{n} \frac{(t_i\mu(X_i,\beta))^{y_i}}{y_{i!}} \quad (1.22)$$

Taking the natural log of the likelihood

$$LnL(y,\beta) = -\sum_{i=1}^{n} y \ln(t_i\mu(X_i,\beta) + \sum_{i=1}^{n} y \ln(t_i\mu(X_i,\beta)) + constant \text{independent of} \beta \quad (1.23)$$

$$LnL(y,\beta) = \sum_{i=1}^{n} y \ln(t_i\mu(X_i,\beta)) - \sum_{i=1}^{n} y \ln(t_i\mu(X_i,\beta) \quad (1.24)$$

Taking the partial derivative of the likelihood function and equating to zero we have

$$S_j(\beta) \to \frac{\partial \ln l(y,\beta)}{\partial \beta_j} \to \sum_{i=1}^{n} y \frac{e^{Xi^t\beta}}{e^{Xi^t\beta}} x_{ij} - \sum_{i=1}^{n} t_i e^{Xi^t\beta} x_{ij} = 0 \quad (1.25)$$

$$S_j(\beta) \to \sum_{i=1}^{n} (y - t_i e^{Xi^t\beta} x_{ij}) = 0 \quad (1.26)$$

$$\hat{S}_j(\beta) = \begin{pmatrix} S_1(\hat{\beta}) \\ \vdots \\ S_p(\hat{\beta}) \end{pmatrix} = X^t(\hat{Y} - \hat{\mu}) \quad (1.27)$$

$$\hat{\mu} = \begin{pmatrix} t_i e^{Xi^t\beta} \\ \vdots \\ t_i e^{Xi^t\beta} \end{pmatrix} \quad (1.28)$$

Equations (1.27) and (1.28) are called score equations.

2.4 FORMULATION OF THE FISHER INFORMATION OF POISSON REGRESSION

From Equation (1.22)

$$\frac{\partial \ln l(y,\beta)}{\partial \beta_j} = \sum_{i=1}^{n} (y - t_i e^{Xi^t\beta} x_{ij}) \quad (1.29)$$

$$\frac{\partial^2 \ln(y,\beta)}{\partial\beta_r\beta_s} = \frac{\partial}{\partial\beta_r}\left((y - t_i e^{Xi^t\beta} x_{ij)}\right) \quad (1.30)$$

$$\frac{\partial^2 \ln(y,\beta)}{\partial\beta_r\beta_s} = -\sum_{i=1}^{n} t_i e^{Xi^t\beta}\, x_{is}x_{ir} \quad (1.31)$$

$$\text{IRS} \rightarrow E\left(-\frac{\partial^2 \ln(y,\beta)}{\partial\beta_r\beta_s}\right) = \sum_{i=1}^{n} t_i e^{Xi^t\beta}\, x_{is}x_{ir} \quad (1.32)$$

$$\rightarrow (\beta) \rightarrow (irs)r,_s = 1,2,\ldots,p = X^t D(\beta)\, X$$

Where $D(\beta) = diag$

2.5 MODELS VALIDATION

The models was validated using R package on maternal deaths data obtained at the Bio-Statistics Department of the Obstetrics & Gynecology directorate of the State Specialists Hospital (OSSH) in Akure, Nigeria for the period of January 2005 to December 2014.

2.6 RESULTS AND DISCUSSIONS

Analysis carried out on the data shows that the mean incidence of maternal deaths was high especially in years 2009 and 2013 with 0.79 and 0.75 percent respectively.

In assessing the goodness of fit for incidence of maternal mortality model, R statistical package was used to fit the data and the results are presented in Table 1 below.

Table 1: Criteria for Assessing Goodness of Fit for Incidence of Maternal Mortality Model

Criteria for assessing Goodness of fit			
Criterion	DF	Value	Value/DF
Deviance	99	98.7236	1.0028
Scaled Deviance	99	98.7236	1.0028
Pearson Chi-square	99	98.6645	1.0034
Scaled Person chi-square	99	98.6645	1.0034
Log likelihood		292.2932	

Table 1 above showed that the dispersion parameter (DP) is 1.0028, indicating a clear absence of over dispersion in the data. Hence, Poisson model is an appropriate model.

The result of the parameter estimates of the maternal mortality using the new proposed Poisson regression model is presented in table 2 below.

Table 2: Parameter estimate of incidence of maternal mortality

Parameter	**Time**	**DF**	**Estimate**	**Std.Error**	**Wald**	**95% C.I**	**Chi-square**	**P-value**
Intercept		1	2.0790	0.3536	1.386	2.772	34.593	0.001
Year	2005	1	0.1180	0.4859	-0.835	1.070	0.059	0.808
Year	2006	1	-0.470	0.5701	-1.587	0.647	0.680	0.410

Year	2007	1	0.318	0.4647	-0.592	1.229	0.470	0.493
Year	2008	1	0.348	0.5000	-0.980	0.980	0.000	1.000
Year	2009	1	-2.079	1.0607	-4.158	-.001	3.844	0.050
Year	2010	1	-0.134	0.5175	-1.148	0.881	0.067	0.796
Year	2011	1	0.223	0.4743	-0.707	1.153	.221	0.638
Year	2012	1	0.223	0.4748	-0.707	1.156	0.224	0.636
Year	2013	1	0.225	0.4787	-0.709	0.141	0.221	0.151
Year	2014	0	0	0	0	0		
Scale		0	1	0	1	1		

From the result obtained above, the linear form or the model of the significant maternal death can be written in the form

$$\widehat{Y_5} = 2.0790 - 2.079X_5$$

Table 2 shows that there was a statistically significant maternal mortality incidence in the year 2009 relative to year 2014, the chi-square value is 3.844 with p-values of 0.050. Hence compared to year 2014, incidence of maternal death was significantly high in 2009. The changes in the signs of the estimated values of the parameter shows that there is no constant rate of incidence of maternal deaths for the ten year period. That is there is a year to year variation in the occurrence of incidence of maternal death. Finally, since all the p- values are greater than 0.05 except for year 2009, there is statistical evidence that the model fits the data appropriately.

3.0 CONCLUSION

This study showed that Poisson regression model is the most suitable for analyzing count data as far as the assumptions of the model are satisfied. As much as Poisson regression model is a suitable model for analyzing count data, the major challenge may be the problem of over-dispersion or under-dispersion of data set which was adequately taken care of in our data set. It can also be concluded from the analysis that within ten (10) years period, years 2009 and 2013 recorded the highest deaths of 17 and 15 respectively. From the estimated Poisson regression model, changes in the sign of the parameters from year 2005 to 2013 shows that the mean incidence of maternal deaths at the hospital changes from year to year compared to 2014 which happened to be the referenced year. The results also showed that there is statistically significant maternal mortality incidence in 2009.

This study recommend that, to solve the problem of over-dispersion or under-dispersion which may occur in any count data set, other extension of Poisson regression model such as negative Binomial model and the excess zero inflated models should be utilized.

REFERENCES

Agrestin A. (2007). An Introduction to Categorical Data Analysis, 2nd Edition, John Wiley and Sons; Hoboken, New Jersey.

Aitkin, M., Anderson, D., Francis, B. & Hinde, J. (1990). Statistical Modelling in GLIM, Oxford University Press, New York.

Bucher DJ, Li SS, Kehoe JM & Kibourne ED (1976). "Chromatographic Isolation of the Hemagglutinin Polypeptides from Influenza Virus Vaccine and Determination of their Amino-terminal Sequences". Proc Natl Acad Sci U S A. 1976 Jan; 73(1): 238–242. PMCID: PMC335876.

Fisher, R. A., Thornton, H. G. & Mackenzie, W. A. (1922). The accuracy of the plating method of estimating the density of bacterial populations. Annals of Applied Biology 325-359. [CP22 in Bennett (1971), vol. 1.].

Nelder, J .A. & Wedderburn R. W. M. (1972). Generalized linear models. Journal of the statistical society. p370–384.

Noriszura I & Abdul A. J. (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. Casualty Actuarial Society Forum, Winter.

Renshaw, A. E. (1994). "Modelling the Claims Process in the Presence of Covariates", ASTIN Bulletin Vol. 24, No. 2, 265-285.

Smart A. Sarpony & Abena K. Brobbey (2013). "Poisson Regression Modeling for Incidence of Maternal Deaths in Ghana". Mathematical Theory and Modeling ISSN 2225-0522 (online) Vol., 3, No. 2.

World Health Organization (WHO), UNICEF, UNFPA and the World Bank (2007). Maternal mortality in 2005: Estimates developed by WHO, UNICEF, UNFPA and The World Bank. Geneva: World Health Organization

WHO, UNICEF, UNFPA and The World Bank (2015). *Trends in Maternal Mortality: 1990 to 2015*, WHO, Geneva.

Yunos, Z. M., Shamsuddin, S. M., Alwee, R., Ismail, N., & Salleh@Sallehuddin, R. (2016). Comparative performance for predictive modelling in motor insurance claims. *Jurnal Teknologi*, *78*(12-2), 29-34. DOI: 10.11113/jt.v78.10118

Zamani, H., Faroughi, P., & Ismail, N. (2016). Bivariate generalized Poisson regression model: applications on health care data. Empirical Economics, 1-15. DOI: 10.1007/s00181-015-1051-7

Zamani, H., Faroughi, P., & Ismail, N. (2014). Estimation of count data using mixed Poisson, generalized Poisson and finite Poisson mixture regression models. In *AIP Conference Proceedings* (Vol. 1602, pp. 1144-1150). American Institute of Physics Inc.. DOI: 10.1063/1.4882628