

Comparison between Quasi-Poisson and Negative Binomial in Handling Over-Dispersion: A Study of Risk Factors Associated with Malaria among Children Younger than Five Years.

Olateju A. Bamigbala, B.Sc.*; Adesupo A. Akinrefon, M.Sc.;
and Oludele R. Afolabi, M.Tech.

Department of Statistics and Operations Research, Modibbo Adama University of Technology,
Yola Adamawa State, Nigeria.

E-mail: olatejubamigbala@yahoo.com*
akinrefon.adesupo@mautech.edu.ng

ABSTRACT

This study used Poisson Regression model, while the performance of Quasi-Poisson Model and Negative Binomial Regression in handling over-dispersion was compared. The variables on which the data were collected include gender, age group, year, and hospital type. By selecting the model of best fit, the AIC of model 3 was found to be 376.67 and this model examines the interaction between hospital type and year of occurrence independent of age of patient, year, and their gender. The study revealed that the odds that female children will suffer from malaria is about 1.0465 times that of males (about 5% more than males). Children under 1 year seem to be the most hit from malaria than those between 1-3 years and 4-5 years with odds of 0.5697 and 0.4583 times as likely to have malaria as those under 1 year. The study further revealed year 2013 compared to 2014 and 2015, the odds that the infant suffers from malaria are 0.2209 and 0.0787 which are less than 1 indicating that babies with malaria decreases every year.

(Keywords: malaria, quasi-Poisson, deviance, negative binomial, goodness of fit)

INTRODUCTION

In modeling count data using Poisson regression models, a commonly experienced problem is that of over-dispersion (where the variance of the response variable exceeds its mean). The over-dispersion downsides will also have consequences on the value of the standard error estimators for smaller, which can further lead to inferential errors in the parameters (Melliana *et al.*, 2013).

Quasi-Poisson and Negative Binomial regression models have equal numbers of parameters, and either could be used for over-dispersed count data. While they often give similar results, there can be striking differences in estimating the effects of covariates. The variance of a Quasi-Poisson model is a linear function of the mean while the variance of a negative binomial model is a quadratic function of the mean (Ver Hoef, and Boveng, 2007). A common way to deal with over-dispersion for counts is to use a generalized linear model framework (McCullagh and Nelder, 1989), where the most common approach is a “quasi-likelihood,” with Poisson-like assumptions (also called the Quasi-Poisson), or a Negative Binomial model.

Globally, approximately 3.28 billion people are at risk of being infected with malaria and most cases occurring in Sub-Sahara Africa. Nigeria has the highest prevalence of falciparum malaria compared to any other country in the world, with 300,000 Nigerians are dying each year from this disease (WHO, 2010). Pregnant women and infants with malaria are high risk groups. In pregnancy, falciparum malaria is associated with maternal anemia and results in subsequent stillbirths, low birth weights, and can also lead to an increased risk in infant mortality.

The adverse effects of falciparum malaria on the fetoplacental unit are particularly apparent in women in the third trimester (Okoko *et al.*, 2002; Bassiouny and Al-Maktari, 2005). In fact, malaria in pregnancy has been linked to estimates of between 0.4 to 1.7 million deaths annually in Sub-Sahara Africa alone (Murphy and Brenan, 2001). Complications from malaria occur due to mismanagement of the treatment of the disease coupled with the lack of prompt and effective treatment (Ruebush *et al.*, 1995).

Moreover, the death of over eighty percent of children inflicted with malaria occurs within the first 48 hours of infection (Dzeing- Ella *et al.*, 2005). Adamawa State, Northeast Nigeria, is highly endemic for the *Plasmodium falciparum* parasite (NBS, 2009) and there are sporadic outbreaks of the less severe form of malaria caused by *P. malariae*. Despite the fact that *P. malariae* infections (also known as quartan fever due to the 3 day periodic cycles of infection) are mildly acute in the early stages, if untreated chronic infections with this form of malaria will result in nephrosis which can be fatal (Carter and Mendis, 2002; Collins and Jeffrey, 2007). Childhood morbidity is very high in Adamawa (NBS, 2009); to reduce this, prompt and accurate diagnosis to the appropriate treatment of malaria is essential.

According to Wenceslaus (2000) the malaria situation in Sub-Saharan Africa is grim and the disease now constitutes a leading cause of poverty in the region. This is because Sub-Sahara African region has the greatest number of people exposed to malaria transmission, greatest burden of malaria morbidity and mortality in the world (WHO, 1996). The problems associated with malaria treatment in Africa had substantially increased the rates of illness and death (Peter *et al.*, 2000). Furthermore, in these areas, most cases of severe malaria occur among children aged between one and three years of age.

Therefore, against this background, this study intends to assess the influence of risk factors on morbidity counts among infants infected with malaria.

The analysis of Poisson regression in this study begins with the formulation of the model. It is then followed by the estimation of parameters, after the estimates are obtained. The goodness-of-fit of the model is checked, this is done by using Pearson chi-square and deviance statistic. After that, test for over dispersion is done. For cases of over-dispersion using Quasi-Poisson model and Negative Binomial regression was be used to overcome this.

METHODOLOGY

The data for this work were obtained from Specialist Hospital and Valli Clinic, Yola of Adamawa State, Nigeria spanning between 2013

and 2015. The data includes variables on age, gender, hospital type and year of occurrence.

Poisson Regression Model

Poisson regression is a non-linear regression analysis of the Poisson distribution, where the analysis is highly suitable for use in analyzing discrete data (count), if the mean equal the variance process. In Poisson regression it is assumed that the dependent variable Y , number of occurrence of an event, has a Poisson distribution given the independent variables X_1, X_2, \dots, X_p .

$$\Pr(Y=y) = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2,$$

Where $E(Y) = \mu$ and $\text{Var}(Y) = \mu$. This is called the equi-dispersion property of the Poisson distribution.

The log of the mean μ is assumed to be a linear function of the independent variables, that is,

$$\ln \mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

Where $Y \sim P(\mu)$

or equivalently,

$$\mu = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}$$

This is the model for analyzing count data.

Sometimes, the response may be in the form of events of certain type that occur over time, space or some other index of size. In this situation, it is often relevant to model the data as the rate at which events occur. When a response count Y has index (such as population size) equal to t , the sample rate of occurrence is Y/t . The expected value for rate is μ/t . Thus, for analysis rate data, the model can be written as:

$$\ln\left(\frac{\mu}{t}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

This model has an equivalent representation as,

$$In\mu - Int = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

The adjustment term $-Int$, on the left-hand side of the equation is called an *offset*.

Over-Dispersion Handling

Over-dispersion (also called extra-Poisson variation) occurs if $Var(Y) > \mu$. If $Var(Y) < \mu$, the problem is called under-dispersion. Under-dispersion seldom occurs in the analysis.

Estimating the Over-Dispersion Parameter

Dispersion parameter, ϕ can be estimated based only on the first two moment of Y . from the relationship:

$$E\left(\frac{(Y_i - \mu_i)^2}{\mu_i}\right) = \phi$$

$$\hat{\phi} = \frac{1}{(N-p)} \sum_{i=1}^N \frac{(Y_i - \hat{\mu})^2}{\hat{\mu}_i}$$

$\hat{\phi}$ is the scaled Pearson Chi-Square, the scaled Deviance (Deviance Statistic divided by its degree of freedom). Other than scaled Pearson Chi-Square, the Scaled Deviance (Deviance Statistic divided by its degree of freedom) is also common in estimating the dispersion parameter. This is given by:

$$\hat{\phi} = \frac{1}{(N-p)} \left[2 \sum_{i=1}^N Y_i \ln \left(\frac{Y_i}{\hat{\mu}_i} \right) \right]$$

From the equation $Var(Y) = \phi\mu$, it is clear that if there is no over-dispersion, the estimated $\hat{\phi}$ will be close to 1.

Testing for Over-Dispersion

Even though over-dispersion parameter is obtained as above, one still need to test for over-

dispersion in order to see whether or not over-dispersion exists. To do so, the following hypotheses are tested.

Model Checking Using Pearson Chi-Squares and Deviance: If the values for both Pearson Chi-squares and deviance are close to the degree of freedom, $N-p$, the model may be considered as adequate. To check the goodness of fit of the model, the following hypotheses are require.

H_0 : The model has a good fit

versus

H_1 : The model lacks good fit

Negative Binomial Regression Analysis: The negative binomial regression model is derived by rewriting Poisson regression model such that:

$$In \mu = \beta_0 + \beta_i X_i + \varepsilon_i$$

The negative Binomial regression distribution has the form.

$$P(Y = y) = \frac{\Gamma(1/\alpha + y)}{\Gamma(1/\alpha) y!} \left[\frac{1/\alpha}{(1/\alpha) + \mu} \right]^{1/\alpha} \left[\frac{\mu}{(1/\alpha) + \mu} \right]^y$$

Where $\Gamma(\cdot)$ is a gamma function. This results in the likelihood function.

$$L(Y_i) = \prod_{i=1}^N \frac{\Gamma(1/\alpha + y)}{\Gamma(1/\alpha) y!} \left[\frac{1/\alpha}{(1/\alpha) + \mu} \right]^{1/\alpha} \left[\frac{\mu}{(1/\alpha) + \mu} \right]^y$$

Maximum likelihood estimation is used to estimate parameters in negative binomial. In addition, the interpretation of regression coefficients for negative binomial regression is the same as for Poisson regression.

RESULTS

The number of infants infected with malaria was modelled using Poisson regression and the results are presented in the Table 1 below. The table contains various models generated and their AIC's.

Table 1 shows that the best model which fit the infants with malaria from 2013 to 2015 is Model 3 because it has the smallest AIC. The AIC of Model 3 was found to be 376.67. The Model 3 examines the interaction between hospital type and year of occurrence independent of age of patient and their gender.

Table 1: The Poisson Regression Models for the Numbers of Infant Infected with Malaria from 2013 to 2015 with their AIC's.

Model	Possible Models Generated	Equations	AIC's
1.	Main Effect Only	$In(y) = \alpha + \beta_1 Age + \beta_2 Gender + \beta_3 Hosp + \beta_4 Year$	397.03
2.	Possible Interaction	$In(y) = \alpha + \beta_1 Age + \beta_2 Gender + \beta_3 Hosp + \beta_4 Year + \beta_5 (Age * Gender)$	398.67
3		$In(y) = \alpha + \beta_1 Age + \beta_2 Gender + \beta_3 Hosp + \beta_4 Year + \beta_5 (Hosp * Year)$	376.67
4		$In(y) = \alpha + \beta_1 Age + \beta_2 Gender + \beta_3 Hosp + \beta_4 Year + \beta_5 (Age * Year)$	399.10
5.	Saturated Model	$In(y) = \alpha + \beta_1 Age + \beta_2 Gender + \beta_3 Hosp + \beta_4 Year + \beta_5 (Age * Gender) + \beta_6 (Age * Year) + \beta_7 (Gender * Year) + \beta_8 (Age * Gender * Year)$	424.50

Table 2: The Parameter estimate of Selected Poisson Regression Model for Numbers of Infants Infected with Malaria from 2013 to 2015.

Predictor	Estimate	Standard Error	Z-value	P-Value	Odd Ratio
Intercept	-2.691169	0.091764	-29.327	2 e -16***	0.0678
Male (Ref)					
Female	0.009238	0.097025	0.095	0.924146	1.0093
Age <1 Year (Ref)					
1 – 3 year	-0.149489	0.122778	-1.218	0.223393	0.8611
4 – 5 years	-1.285848	0.203007	-6.334	2.39e-10***	0.2764
Valli (Ref)					
Specialist Hosp.	2.117695	0.166561	12.714	2e-16***	8.3120
2013 (Ref)					
2014	-1.569403	0.156925	-10.001	< 2e-16***	0.2082
2015	1.608542	0.176847	-9.096	2e-16***	4.9955
Interaction					
(Specialist*2014)	0.302713	0.234353	1.292	0.196462	1.3535
(Specialist*2015)	-2.119945	0.612573	-3.461	0.000539***	0.1200

Significant code: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

Model Checking Using Pearson Chi-Squares and Deviance

If the values for both Pearson Chi-squares and deviance are close to the degree of freedom, N-p, the model may be considered as adequate. To check the goodness of fit of the model, the following hypotheses require.

H_0 : The model has a good fit

H_1 : The model lacks good fit

With all factors constants the odds that babies suffer from malaria is about 0.0678 (7% of babies born experience this). Babies between 4-5 years are 0.2764 times likely to have malaria as babies < 1 years, babies < 1 years are more susceptible to malaria, also, hospital effect cannot be ruled out, the state owned hospital record higher cases of malaria admission etc.

Table 3 shows the Pearson Chi-square of 407.811 and deviance of 236.5239 both on 27 degree of freedom are greater than the critical value of 40.113 indicating lack of good fit. The dispersion parameter was found to be 15.1041 indicating that the model is significant. However, the assumption of equal variance to the mean in Poisson distribution has been violated since the dispersion parameter is not approximately equal to 1.

The dispersion parameter of the above model is 15.0141 which is far greater than 1, an indication of over dispersion in the data. This means that the parameters of the model have been overestimated and the standard errors have been under estimated which will not give a true reflection of the model which could provide appropriate mean number of infants with malaria from 2013 to 2015. To address this error, Quasipoisson model and Negative Binomial regression was used to modify the model to nullify the effect of over dispersion in the data and the result is shown in the Table 4 below.

Table 3: Number of Infants infected with Malaria (Categorical) Using Poisson Regression.

Pearson Chi-Square	407.811 on 27 degree of Freedom
Deviance	236.5239 on 27 degree of Freedom
Dispersion Parameter	15.1041

Table 4: The Parameter Estimate of Using Quasipoisson Model and Negative Binomial Regression for Numbers of Infant Infected with Malaria.

	Quasipoisson Model			Negative Binomial Regression Model		
	Estimated Coefficient	Standard Error	P-Value	Estimated Coefficient	Standard Error	P-Value
Intercept	-2.691169	0.356633	4.07e-08*	-1.64805	0.46887	0.00044*
Male (Ref)						
Female	0.009238	0.377080	0.98063	0.04544	0.35369	0.89778
Age <1 Year (Ref)						
1 – 3 year	-0.149489	0.477164	0.75647	-0.56273	0.49868	0.25914
4 – 5 years	-1.285848	0.788967	0.11476	-0.78013	0.74518	0.29514
Valli (Ref)						
Specialist Hosp.	2.117695	0.647324	0.00292*	1.25204	0.74522	0.09294**
2013 (Ref)						
2014	-1.569403	0.609873	0.01588*	-1.51016	0.57437	0.00856*
2015	-1.608542	0.687300	0.02690*	-2.54242	0.58257	1.28e-05*
Interaction						
Specialist*2014	0.302713	0.910791	0.74218	0.42312	0.80836	0.60068
Specialist*2015	-2.119945	2.380706	0.38108	-1.33682	1.00558	0.18372
Dispersion Parameter	15.10411			Dispersion Parameter	1.133	

* Significant at 5%

** Significant at 10%

The standard errors for the Negative Binomial are lower than that of the Quasi-Poisson Model. The dispersion parameter of Negative Binomial is approximately 1. Negative Binomial handle better accounts for the Extra-Poisson variance parameter than Quasi-Poisson Model.

From the results presented in Table 4, Quasipoisson model maintain the same dispersion parameter with Poisson regression model but these are changes the standard error and reduced number of significance value compared to Poisson regression model as shown in Table 4. It can be seen that the Negative Binomial regression model best fits the data on the numbers of infant infected with malaria because the dispersion parameter reduced from 15.1041 which was giving by the Poisson regression model and Quasi-Poisson model to 1.133.

Interpretation of Coefficient

It is noted that the risk factors year 2014 and 2015 are significant at $\alpha=0.05$ with their respective p-values equal to 0.00856 and 1.28 e-05 while specialist hospital is significant at $\alpha=0.10$ with p-value 0.09294.

The odds that female children will suffer from malaria is about $e^{0.04544} = 1.0465$ times that of males about 5% more than males; children under 1 year seem to be the most hit from malaria than those between 1-3 years and 4-5 years with odds of $e^{-0.56273} = 0.5697$ and $e^{-0.78013} = 0.4583$ times as likely to have malaria as those under 1 year. Clearly, the chance of being affected with malaria reduces as the babies advance in age.

Valli Clinic compared to Specialist Hospital has an odds $e^{1.25204} = 3.4975$, which indicate that in Specialist Hospital are about 3 times as likely to record cases of malaria than Valli Clinic. For year 2013 compared to 2014 and 2015, the odds that infant suffer from malaria are $e^{-1.51016} = 0.2209$ and $e^{-2.54242} = 0.0787$ which are less than 1 indicating that malaria prevalence among infants seem to be decreasing in Adamawa State.

CONCLUSION

In the light of these findings, we suggest that educating mothers about the mechanisms of environmentally related disease transmission, promoting the safe disposal of children's and encouraging safe water storage, also mosquito treated net by NGOs and the government should be distributed to the inhabitants of Yola and its environs. From the result at this study, we infer that there exist hospital effect in dealing with the prevalence of malaria among infants. Some hospitals seem better equipped to handle such cases hence the number of patients recorded. Also, there seem to be a good and vigorous campaign against malaria incidence in infants. This is evident in the steady decline in odds of malaria occurrence per year.

REFERENCES

1. Bassiouny, H.K. and M.T. Al-Maktari. 2005. "Malaria in Late Pregnancy in Al Hodeidah Governorate, Yemen." *East Mediterr Health J.* 11(4):606-17.
2. Carter, R.C. and N.M. Kamini. 2002. "Evolutionary and Historical Aspects of the Burden of Malaria". *Clin Microbiol Review.* 15(4):564-594.
3. Collins, W.E. and G.M. Jeffery. 2007. "Plasmodium malariae: Parasite and Disease". *Clin Microbiol Rev.* 20(4):579-592.
4. Dzeing- Ella, A., P.C.N. Obiang, R. Tehoua, T. Planche, B. Mboza, M. Mbounja, U. Muller-Roemer, J. Jarvis, E. Kendjo, E. Ngou-Milama, P.G. Kremsner, S. Krishna, and M. Kombila. 2005. "Severe Falciparum Malaria in Gabonese Children: Clinical and Laboratory Features". *Malaria Journal.* 4:19.
5. McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models. 2nd Edition.* Chapman and Hall: New York, NY.
6. Melliana, A., Y. Setyorini, H. Eko, S. Rosi, and S. Purhadi. 2013. "The Comparison of Generalized Poisson Regression and Negative Binomial Regression Methods in Overcoming Over-Dispersion". *International Journal of Scientific and Technology Research.* 2(8):255-258.
7. Murphy, S.C. and J.G. Brenan. 2001. "Gaps in the Childhood Malaria Burden in Africa: Cerebral Malaria, Neurological Sequelae, Anemia, Respiratory Distress, Hypoglycemia, and Complications of Pregnancy". *Am. J. Trop. Med. Hyg.* 64(1, 2): 57-67.

8. National Bureau of Statistics (NBS). 2009. "Social Statistics in Nigeria 2009". Federal Republic of Nigeria. Available at http://www.nigerianstat.gov.ng/ext/latest_release/ssd09.pdf.
9. Okoko, B.J., M.O. Ota, L.K. Yamuah, D. Idiong, S.N. Mkpanam, A. Avieka, W.A. Banya, and K. Osinusi. 2002. "Influence of Placental Malaria Infection on Foetal Outcome in the Gambia: Twenty Years after Ian McGregor". *J Health Popul Nutr.* 20(1):4-11.
10. Peter, B.B., M.E. Maryon, and M. Sylvia. 2000. "Combination Therapy for Malaria in Africa". *Bulletin of the World Health Organization.* 78(12): 1377 – 1386.
11. Ruebush, T.K., M.K. Kern, C.C. Campbell, and A.J. Oloo. 1995. "Self-Treatment of Malaria in a Rural Area of Western Kenya". *Bull World Health Organ.* 73(2): 229–236.
12. Ver Hoef, J.M. and P.L. Boveng. 2007. "Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Over-dispersed Count Data?". Publications, Agencies and Staff of the U.S. Department of Commerce: Washington, DC. Paper 142.<http://digitalcommons.unl.edu/usdeptcommercepub/142>
13. Wenceslaus, L.K. 2000. "Roll Back Malaria in Sub-Saharan Africa. Bulletin of the World Health Organization". 78(12):1452 – 1453.
14. WHO (World Health Organization). 2000. "Severe and Complicated Malaria". *Trans. Res. Soc. Trop. Med. Hyg.* 94(suppl).
15. WHO (World Health Organization). 1996. "World Malaria Situation in 1993, Part 1". *World Health Organization. Weekly Epidemiological Record.* 71: 17 – 22.
16. WHO (World Health Organization). 2010. "World Malaria Report 2010". Available at: http://www.who.int/malaria/world_malaria_report_2010/en/index.html

ABOUT THE AUTHORS

Olateju A. Bamgbala, is a Graduate of Statistics from the Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola. He holds a Bachelor of Technology Degree in Statistics.

Adesupo Akinrefon, is a Lecturer in the Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola. He holds a Master of Science

Degree in Statistics, (Ph.D. in view) from the University of Ilorin, Nigeria. His research interests include categorical data analysis, stochastic processes, and modelling.

Oludele R. Afolabi, is a Graduate of Operations Research from the Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola Adamawa State. He holds a Master of Technology degree in Operations Research. His research interest criteria include, multi criteria decision analysis, time series analysis, and modelling.

SUGGESTED CITATION

Bamigbala, O.A., A.A. Akinrefon, and O.R. Afolabi. 2016. "Comparison between Quasi-Poisson and Negative Binomial in Handling Over-Dispersion: A Study of Risk Factors Associated with Malaria among Children Younger than Five Years". *Pacific Journal of Science and Technology.* 17(2):293-299.

