Modeling Length of Stay in Hospital Using Generalized Linear Models: A Case Study at Tamale Teaching Hospital

A.R. Abdul-Aziz
Lecturer
Department of Mathematics and Statistics
Kumasi Polytechnic, Kumasi, Ghana

L. Munyakazi
Associate Professor and Head
Department of Mathematics and Statistics, Ghana

N.N.N. Nsowah-Nuamah
Professor in Statistics and Rector of the Kumasi Polytechnic, Ghana

Abstract

Malaria remains a major health problem of our time. The disease plagues human society and impacts obnoxious and immeasurable burden on human population. Since children particularly are the worse affected, it is pertinent to examine some of the demographic factors connections in malaria incidence. This paper is aimed at modeling length of stay using covariates from hospital register and also assessing interaction effects. Covariates were assessed through morbidity and mortality register at the Tamale Teaching hospital. The results showed that expected length of stay per day at hospital was influenced by covariates such as outcome on admission, referral status, distance, treatment, season and sex of children administered as malaria patients at the Tamale Teaching hospital. Also, there was sufficient evidence to show that there is interaction effect between; outcome on admission and referral status, sex and referral status and treatment and the season in which patients were admitted as malaria patients.

Keywords: Malaria, Patient, Length, Poisson, Negative Binomial and Covariates

1. Introduction

Malaria remains a major health problem of our time. The disease plagues human society and impacts obnoxious and immeasurable burden on human population. It displays full explosive power of vector-borne infections, erupting suddenness and intensity that can overwhelm vulnerable communities (Kiszewski and Teklehaimanot, 2004). Effective anti-malarial control strategies for young children in Africa must urgently be improved, as this group is at highest risk of morbidity and mortality due to *Plasmodium falciparum* malaria (Snow et al., 2005 and Hay et al., 2004). In most parts of sub-Saharan Africa, individual control measures concentrate on the reduction of parasite transmission through insecticide treated bed nets and treatment of acute malaria episodes (Breman et al., 2004), the latter often being inappropriate due to increasing parasite drug resistance (May et al., 2003 and Plowe et al., 2005).

Worldwide, more than 500 million malaria attacks occur every year, and about 2 million people die of *Plasmodium falciparum* malaria. Sub-Saharan Africa carries most of the burden, and in regions of stable transmission children less than 5 years of age are at highest risk of malaria morbidity and mortality (Snow et al. 2005). Given the magnitude of the problem on the continent, exact targeting of malaria-control measures is needed for cost-effective application of proven and/or new interventions, because it has been shown that only 20% of the population at risk suffers 80% of all infections (Smith et al. 2005 and Woolhouse et al. 1997). Moreover, malaria risk varies markedly across Africa and, most importantly, within countries.

For example, a number of studies have shown that malaria vector distribution, transmission rates, and incidence can vary widely over short distances, between neighboring villages, and even within a single settlement as a result of small-area variations in risk factors (Greenwood et al. 1989 and Lindsay et al. 1990). The identification and understanding of this variation is important, because it allows the detection of high risk groups and selective targeting of intervention measures (Carter et al. 200). The burden of malaria is so enormous and no tolerable limit needs to be admitted. It is astonishing that since the dawn of the pernicious disease about 3200 – 7000 years ago, despite the global efforts, malaria is still claiming it tolls. Its burden on the health of the people is alarming. As it has been mentioned, it is a leading cause of mortality especially in children. It may lead to low birth weight which is as a result of incessant malaria during pregnancy and is a major cause of death in the first month of life. Repeated infection in children may lead to cerebral malaria or severe anaemia which may result in mortality in no time, and increase susceptibility to other childhood illnesses. An estimated 2% of children who survive form malaria infections affecting the brain (Cerebral malaria) suffer from learning impairments and disabilities due to brain damage, including epilepsy and spasticity (Murphy *et al.*, 2001).

This indicates that malaria may also lead to longterm disability. In all malaria-endemic countries in Africa 25-40% (average 30%) of all outpatient clinic visits are for malaria and between 20% - 50% of all hospital admissions are a consequence of malaria, (WHO, 2003). In Nigeria, over 50% of outpatients' attendances and 40% of hospital admission, 30% of child mortality and 10% of maternal mortality are due to malaria (Mosanya 2000 and Abdulkareem, 2001). In Cote D'ivoire, up to 40% of outpatient attendance, over 50% of hospital admissions and 20% of hospital deaths are due to malaria. The data are similar in most endemic countries.

Malaria control has been a major challenge in Ghana, especially in Northern Ghana where it has constituted a major setback for the children growth and development. The burden of malaria is geographically biased to the Ghanaian community. It is thus a health problem that has attracted concerted efforts over the years. It is quite unfortunate that despite these efforts malaria still remains a leading cause of morbidity and mortality especially in Northern Ghana. Since it is a problem that affects the people and children particularly, it is pertinent to examine some of the demographic factors connections in malaria incidence. It is in line with this that the paper presented here is aimed at modeling length of stay using covariates from hospital register and also assessing interaction effects. Covariates were assessed through morbidity and mortality register information systems.

The incidence density of infection and disease caused by *Plasmodium falciparum* in children aged six to 24 months living in the holoendemic Sahel of northern Ghana was measured during the wet and dry seasons of 1996 and 1997 (Baird et al., 2002). Malaria has been well controlled or eliminated in the five northernmost African countries: Algeria, Egypt, Libya, Morocco and Tunisia (WHO, 2003). The South of the Sahara is the most endemic region in the world and it is in the region that malaria claims its highest toll. Malaria becomes the most pernicious and prevalent health problem in the region where transmission occurs. It is responsible for at least a million deaths each year with Africa bearing the brunt of the disease accounting for more than 90% of the whole cases (Wellcome, 2002; WHO, 2003; White, 2004). Every year about 300 million clinical cases are also reported (Bloland et al., 2000; Nuwaha, 2001; Moree and Ewart, 2004; Breman et al., 2004; Barat et al., 2004; Agyepong and Kankeye- Kayonda, 2004). This indicates that malaria constitutes a major health constraint to the people and hinders them from day to day social activities as a result of clinical attendance, admissions, or ill-health. Malaria is a prime etiologic factor of slower economic growth in Africa as a result of loss of quality manpower; productivity which might be expressed in terms of absenteeism from employment, discounted future life time earning of those who die, lost school days and permanent neurological and other damages associated with *falciparum* malaria. Malaria is a leading cause of child morbidity and mortality in Africa as children are said to account for the 90% of the whole cases (WHO, 1997; Baume et al. 2000; WHO, 2003; NPC, 2004). Kevin et al. (2007) reported that chloroquine has been first-line therapy for vivax malaria since 1946, and the emergence of resistance to the drug further complicates therapeutic management decisions.

In the cases of self-treatment, it is also usually based on presumptive treatment, and this has been implicated in development and spread of antimalarial drug resistance. Non-compliance with therapeutic regimen is common in self-treatment. McCombie (2002) reported that the widespread of stopping medication when symptoms resolve as drug may be saved for future episode is well known. Knowledge of correct dosage varies, in some cases it may be lacking. When it is lacking, it might not be given. This exposes the parasites to sub-optimal drug level and may result in the development of resistance.

2. Methods

2.1. Study Area and Source of Data

Tamale metropolis, one of the 20 districts located in Northern Region, is generally classified as malaria endemic, although the highland zones in the central parts of the district are of low transmission and may be prone to epidemic malaria. Tamale has a total population of 2,468,557 people as per the last 2010 census. Currently the growth rate of Tamale is 2.9% per annum (*Ghana* Statistical Service, 2011). The Tamale Teaching Hospital is a state run teaching and referral hospital built in 1974. It was one of the admired medical complexes in Ghana. It has a capacity of about 380 patient beds; a four -storey structure that houses four wards, the general purpose theatre and an X-ray Unit. The obstetric/gynecology ward; and antenatal Units (Gunu, 2009).

2.2. Data Collection and Data Management

Data used in this study were obtained as primary data from discharge records of all pediatric hospital admissions at Tamale Teaching Hospital, from 1st January 2008 to 31st December 2010. The hospital, with about 380 patient beds, is the largest facility in the metropolis which serves both as the first consultation point for patients within its catchment, and as a referral centre for about other 15 primary health centers. These facilities are managed by the Ministry of Health and he Ghana Health Service with support from other some partners (Gunu, 2009). For this study *the actual length of stay in the hospital due to malaria* is our primary variable of interest. Each case was clinically assessed and definitively confirmed as malaria on admission. The registers included the following collected determinants for *individual patient*: patients' age, gender, date of admission and discharge, outcome (i.e. death, discharged home, or absconded), location of residence, cost for treatment, referral status, and treatment given. The following variables were coded: *outcome* (1 = dead and alive = 0); *season* of the year when admitted (1 = wet season from April to October, 0 = dry season from November to March); the actual *treatment* given to the patient (1 = artesunate amodiaquine, 0 = quinine); the *distance* to the hospital (1 = distance > 5 km, 0 = distance ≤ 5 km). The distance of 5 km was chosen to reflect travel time of 1 hour on foot; and finally the variable that defines whether the patient was *referred* or not. Children who used the hospital as a first point of consultation were given a code of '0' and those referred to the hospital from peripheral health facilities in the metropolis given the code '1'.

2.3 Model Specification, Tests and Hypotheses

The number of days each patient stay at the hospital (in days) is recorded as *count*. As a consequence, the *Poisson regression model* is particularly appropriate for this type of response. The model is non-linear and describes the mean number of days. The distribution of the number of days using a Poisson model is of the form:

$$Prob[Y = y] = \frac{\exp(-\lambda T)(\lambda T)^y}{y!} \quad y = 0, 1, ... \qquad (1)$$

where y represents the observed number of days, γ is a non-linear regression defines by

$$\lambda = exp(\beta' x) > 0 \qquad (2)$$

The model matrix x is a *px1* vector of explanatory variables and β is a *px1* vector of regression parameters.
The conditional mean

$$E[Y|x] = \lambda = exp(\beta' x) \qquad (3)$$

The variance, (also called the heteroskedastic conditional variance) of the random variable is *constrained to be equal to* the mean i.e.,

$$Var[Y|x] = \lambda \qquad (4)$$

In most practical situations such as ours, the parameters in (2), namely the β_i, are estimated by the method of maximum likelihood, although other methods are available (SAS, 2004; Hilbe and Greene (2008); Greene (2003); and Wooldridge (2003). Because of the equidispersion assumption which, in most cases, is atypical, our analysis added a fit of a more general specification model known as the *Negative Binomial* (NB) model. This model (at least one alternative of) has become the standard choice for basic statistical analysis of count data (Greene, 2008) especially among statisticians i.e., the so-called *generalized linear model*. Anscombe (1949) is among the first to use this as a tool for the analysis of count. The Negative Binomial probability distribution of Y is

$$Prob[Y = y] = \left(\frac{r}{r+\lambda}\right)^r \frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)} \left(\frac{\lambda}{r+\lambda}\right)^y \qquad y = 0, 1, ... \qquad (5)$$

where $\Gamma$is the gamma function (SAS, 2007; Kiebel and Holmes (2003). In this case, the mean of the negative binomial distribution is the same as that of the Poisson i.e.,

$$E(y|x] = \lambda \quad (6)$$

but the variance is

$$Var[Y|x] = \lambda + {\lambda^2}/{r} \quad (7)$$

where r is the so-called *dispersion* parameter. The conditional variance of the Negative Binomial distribution exceeds the conditional mean. A general class of Negative Binomial is given in Cameron and Trivedi (1986) and Greene (2008). The variance function is

$$\mu_i + \alpha\mu_i^p \quad (8)$$

which gives rise to two Negative Binomial models: the so-called NEGBIN2 corresponds to p = 2 i.e., the variance function is linear in the mean and NEGBIN1 is for p = 1, i.e., the variance function is linear in the mean (SAS, 2007; Greene, 2008). The results from all variants will be presented in this study alongside the corresponding main 7diagnostic statistics.

From equation (7), the larger value of *r* reduces the NB model to a Poisson model a fact that clearly shows the relationship between the two formulations. In practice, it is often the case that the collected explanatory variables is not exhaustive enough to fully explain the observed length of days the afflicted children spent in hospital. While in classical linear model ε is assumed to follow a normal distribution with the introduction of an *identity* link between $E(Y)$ and μ, the *log* link function is used for Poisson regression model to link the number of days spent in hospital swith the linear combination of parameters, whereas the *log-gamma* link facilitates the relationship in NB models.

2.4 G**oodness of fit Tests**

The Negative Binomial is more general than the Poisson model. In fact it is derived as a gamma mixture of Poisson random variables (SAS, 2007). As a consequence, the Poisson model is *nested* within the Negative Binomial model and therefore a **Likelihood Ratio Test** (LRT) is applicable (Hilbe and Greene, 2008;Greene, 2008; Cameron and Trivedi, 1998). The likelihood ratio statistic is

$$\chi^2[J] = 2(logL_1 - logL_0) \quad (9)$$

where $Log\ L_1$ is the log-likelihood of the full or *unrestricted* estimator and $Log\ L_0$ is the log- likelihood of the *restricted* (Poisson model). In both cases, the same LRT is used to test whether the model (Poisson or Negative Binomial) fits the data significantly better than the null model which contains only the *intercept*. The resulting statistics has a $Chi - Square$ distribution with the number of the degrees-of-freedom equals the difference in degrees-of-freedom of the respective models. In the first case, a *small p-value* implies that the NB describes the data better than the Poisson model, whereas the second case implies that the *linear* combination of predictors fits the data in a satisfactory manner. The null hypothesis

$$H_0: \mathrm{c}(\beta) = 0 \quad (10)$$

is rejected in favor of the alternative. On the other hand, the H_0 is rejected if LRT is greater than Chi-Square $(1 - 2\alpha; 1\,df)$. This approach is analogous to the extra-sum of squares principle or the so-called Principle of Conditional Error) often applied to the analysis of continuous normally distributed data (Draper and Smith, 1998).

For *each* predictor, a Wald statistics is computed as

$$(\beta_j / SE_{\beta_j})^2 \quad (11)$$

also distributed as a Chi-Square with 1 degree-of freedom. Computation of the Lagrange Multiplier (LM) also referred to as the *score test* can be found in SAS (2007) and Hilbe and Greene (2008). Other measures of goodness-of-fit in the Poisson model (Hilbe and Greene, 2008) include the *Deviance* statistic

$$G^2 = 2\sum_{i=1}^{N} y_i \log(\frac{y_i}{\lambda_i}) \quad (13)$$

The degrees-of freedom of the deviance statistic is $n - p$ where n is the number of observations and p represents the number of predictors including the intercept.

This statistic is used in a similar manner as the Likelihood Ratio Test by using the ratio $\frac{Difference\ \ in\ Deviances}{Difference\ \ in\ the\ degrees\ \ of\ \ freedom}$.

The obtained ratio is chi-square distributed with *df* equals to the above difference in *df*. The last criteria this study reports is the *Pearson goodness-of-fit* statistics (Breslow, 1984) defined as

$$C^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} \quad (14)$$

Another practical interpretation is seen from the fact that the expected value of a chi-square random variable is equal its degrees of freedom, therefore the ratio $\frac{deviance}{degree\ \ of freedom}$ should be about 1 to indicate a good fit for both the deviance and the Pearson Chi-square criteria. Large or small values of the ratio may indicate an over-dispersion response or a misspecified model. The later suggests that the analyst should be reasonably sure that the lack of fit is not due to poor specification of the systematic part of the model before an over-dispersion correction is done. Additional measures of goodness of fit include the Akaike Information Criteria (AIC) and the Swhartz Bayesian Criteria (SBC) familiar to users of generalized linear fixed and/or mixed effects models. The implementation of all the above model fits is done through SAS (2007).

3. Analysis and Results

3.1 Descriptive statistics

Test of normality confirms serious departure from normality by all tests below:

The magnitude of skewness was very large (= 4.21) implying the distribution is skewed to the right, that is, more observations are on the left. The measure of thinness of tails i.e., kurtosis is 15.74 indicating a higher peak and thin tails compared to a normal distribution (Hun, 2008).

3.2 Diagnostic Statistics of the models

The Poisson model, the Poisson corrected for overdispersion due to unmeasured heterogeneity that deflates the standard errors, and the Negative Binomial (p=2) models' results are reported in Table 3. The correction affects only the standard errors of the parameter estimates and not the estimates themselves. The deviances and the Pearson Chi-Square are used to assess the existence of overdispersion. The full-log likelihood is implemented to test the hypotheses that all the coefficients are zero by taking the difference in their respective full log-likelihoods and the full likelihood of the restricted model i.e., the model contains only the intercept as specified in (10). The resulting statistics are Chi-Square distributed with the degrees-of-freedom equal to the difference in the corresponding degrees-of-freedom (11 in all cases). Table 3 summarizes the results.

3.3 Inferential Analysis of the Model

From the results of the Negative Binomial, the covariates; *outcome on admission, referral status*, *age*, *sex*, *distance*, *treatment*, the interaction between *referral status* and *outcome* on admission and the interaction between *sex* and *referral status* are *statistically* significant at $\alpha = 0.05$ with their respective p-values equal to 0.000, 0.0001, 0.000, 0.0316, 0.000, 0.0005, 0.000 and 0.0186. Against this backdrop, therefore, these covariates are relevant in predicting length of stay per day due to malaria at the Tamale Teaching hospital. From Table 1 below, it is revealing to note that, the covariates season and interaction between season and treatment are not **statistically** significant at $\alpha = 0.05$, with p-values 0.2172 and 0.0774 respectively.

As in Table 2, the strongest covariate of the length of stay per day of malaria administered patient was *outcome* on admission recording 2.143days. This indicate that the expected length of stay per day increased by 2.143 days for patients on admission who survive as compared to patients who died, controlling for all other covariates in the model. Again, the expected length of stay per day of malaria patients increased by 1.816 for patients who call at the Teaching Hospital as their first point of consulting compared to patients who referred from other peripheral clinics and hospitals when other covariates remain constant. Meanwhile, expected increase in length of stay at the hospital per day of age (in years) equals 1.024. Also, expected increase in length of stay per day is 1.306 for females compared to their male counterparts when other covariates are held constant. Interestingly, it was revealing that, *dry season* increased the expected length of stay per day at the hospital by 1.049 compared to the wet or rainy, holding other covariates constant.

Furthermore, patients who stay within 5 km radius around the hospital reduce the expected length of stay per day by 0.837 compared to patients who travel distances beyond 5 km to the hospital when other covariates are held constant. However, patients who were treated with *quinine* increased the expected length of stay per day at the hospital by 1.175 compared to patients who received *artesunate ammodiaquine* (ACT) treatment.

Moreover, the interaction effect of patients who visited the teaching hospital as first point of consultation and were alive upon admission reduced the expected length of stay per day by 0.515 compared to patients who were referred from other hospitals and were dead upon admission. Again, the expected decrease in the length of stay per day was 0.739 due to the interaction effect of non-referred patients and female patients compared to male patients on referral, controlling for other covariates. It was also observed that, the interaction effect of patients who were treated with quinine in the dry season reduced the expected length of stay per day by 0.894 compared to patients who received ACT treatment during the wet or rainy season, holding other covariates constant.

4. Discussion

This study provides evidence of the covariates which influence length of stay per day of malaria patients among children, up to 14 years, at the Tamale Teaching hospital in the northern region of Ghana. The Negative Binomial model indicates that *outcome* upon admission contributes more, among other factors, in terms of influence on expected length of stay per day. It was also observed that non *referral* children spent more days in hospital. This seems to suggest that delay in the process of being transferred to the Tamale Teaching hospital, increased the severity of the malaria, thereby decreasing the expected length of stay per day of such patients. This could be because most referring health facilities may often be faced with stock-out of effective drugs or may not have prompt access to ambulatory support when needed. This also suggests inadequate care being available at primary facilities, regardless of whether they are distant from the hospital or not. It is also possible that referring hospitals are referring the more severe cases which are expected to have higher mortality case.

Distant villages or areas with ill resourced health centers or none at all suggest problems of access to health care, which does translate into high mortality rate, hence reducing the expected length of stay per . Thus the closer the village is to the teaching hospital, the more advantaged the households are in terms of getting early health care and getting treated and discharge in time. The study showed that patients within 5 km of hospital spent fewer days in hospital than those beyond 5 km, and does reflect the fact that nearness to the hospital improved early access to health care. Although, there is the perception that malaria transmission is more intense in the wet season than the dry season, yet the study showed that malaria patients spent fewer days in dry *season*. Though, in Northern Ghana, *season* surprisingly was not even statistically significant covariate in the model as well as its interaction effect with treatment. This could be attributed to the fact all 3 years were lump together in this analysis, implying that an interaction between year and season could improve the understanding of the 'season' effect. One major shortcoming of using this data is that they only represent those patients who visited the Tamale Teaching hospital. Meanwhile, some malaria treatments occur outside the formal hospital and as such are treated at pharmaceutical shops and homes, and only seek medical attention at the hospital if the illness is perceived to be near fatal.

5. Conclusion

The results showed a linear relationship between length of stay per day at hospital and covariates such as outcome on admission, referral status, distance, treatment, season and sex of children administered as malaria patients at the Tamale Teaching hospital. Also, it was found that there is sufficient evidence to show that there is interaction effect between; outcome on admission and referral status, sex and referral status and treatment and the season in which patients were admitted as malaria patients. The following policy interventions are necessary: Stake holders, particularly the Ghana Health service (GHS) should step up efforts to discourage entirely the use of quinine and keep up the on-going campaign on using ACT for malaria treatment. Also, government should expand health centers in the communities; since this could reduce the distance patience have to travel and enabling patients to receive early treatment, thereby reducing length of stay.

Acknowledgement

Our sincere gratitude goes out to Mr. Issah Mohammed Awal, a biomedical technician, at the Tamale Teaching Hospital who assisted in securing the data used for this study.

References

Abdulkareem, H.A.: MACOA and roll back malaria program in Africa. *A Public Lecture Delivered at Afro-Asiatic Institute,* Vienna, Austria (2001).

Agyepong, I. A., and Kangeye-Kayonda, J.: Providing estimate of malaria burden for health planners in resource-poor countries. *American Journal Tropical Hygiene,* 71(2): 162-167 (2004).

Anscombe F. J (1949). The statistical analysis of insect counts based on the negative binomial distribution. Biometrics 165-173.

Baird, J.K., Wiady, I., Sutanihardja A., et al. (2002). Short report: therapeutic efficacy of chloroquine combined with primaquine against Plasmodium falciparum in northeastern Papua, Indonesia. *Am J Trop Med Hyg* 66:659–660.

Baume C., Helitzer, D. and Kachur, S. P.: Patterns of care for childhood malaria in Zambia *Social Science and Medicine*, 51:1491-1503 (2000).

Barat, L.M., Palmer, N., Basu, S., Worrall, E., Hanson, K. and Mills, A.: Do malaria control interventions reach the poor? A view through the equity lens. *American Journal of Tropical Hygiene,* 71(2): 174- 178 (2004).

Bloland, P.B., Ettling, M. and Meek, S.: Combination Therapy for Malaria in Africa: Hype or Hope? *Bull World Health Organ*, **78:** 1378-1388 (2000).

Breslow, N (1984). Extra-Poisson variation in log-linear models. J. Hygiene Camb., 55, 564-581

Breman, j. g., alilio, m. s. and mills, a.: conquering the intolerable burden of malaria: what's new, what's needed: A summary. *America Journal Tropical Hygiene,* 71(2): 1-15 (2004).

Carter R, Mendis KN, Roberts D. *Spatial targeting of interventions against malaria. Bull World Health Organ 2000;78:1401-11.*

Cameron C., Trivedi P., (1998). Regression Analysis of Count Data. Cambridge University Press, New York.

Drakeley, C., Schellenberg, D., Kihonda, J., et al. (2003). *An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. Trop Med Int Health 8:767-77*

Draper, N. R and Smith H. (1998). Applied Regresion Analysis, 3rd ed

Ghana Statistical Service (2011). *2010 Population and Housing Census. Provisional Results. Summary of Findings*. www.ghana.gov.gh/census/phc2010.pdf

Greene, W (2008). Functiona forms of the negative binomial model for count data. Economic letters 99, 585-590.

Greenwood BM. *The microepidemiology of malaria and its importance to malaria control. Trans R Soc Trop Med Hyg 1989;83(Suppl):25-9.*

Gunu, O.M. (2009). *The Deteriorating State of Tamale Teaching Hospital – Tamale, Ghana. Manassas, VA.* http://knol.google.com/k/the-deteriorating-state-of-tamale-teaching-hospital-tamale-ghana#2009

Hilbe J. M, and W. Greene (2008). Count Response Regression Models. Handbook of Statistics, Vol 27 210-252.

Hilbe, J. (1994), "Log Negative Binomial Regression Using the GENMOD Procedure," in *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.

Hun M P (2008). Univariate Analysis and Normality Test Using SAS, Stata, ans SPSS. University Information Technology Services. Center for Statistical and Mathematical omputing. Indiana University.

Kevin, B. J., Schwartz, E., Hoffman, S. L. (2007). Prevention and Treatment of Vivax Malaria, Current Infectious Disease Reports 9**:**39–46

Kiebel S.J and A.P. Holmes (2003). The General Linear Model. The General Linear model. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function.* Academic Press, 2nd edition.

Kiszewski, A. E. and Teklehaimanot, A.: A review of clinical and epidemiologic burden of epidemic malaria. *American. Journal. of Tropical. Medicine and Hygiene,* **71(2):** 128- 135 (2004).

Liao T.F (1999). Interpreting Probabily models. Logit, Probit, and other Generalized Linear Models. Sage University papers. Series: Quantitative Applications in Social Sciences 87 p.

Lindsay SW, Campbell H, Adiamah JH, Greenwood AM, Bangali JE, Greenwood BM. *Malaria in a peri-urban area of The Gambia. Ann Trop Med Parasitol 1990;84:553-62.*

McCombie, S. C.: Self-treatment for malaria: the Evidence and methodological issues. *Health Policy and Planning,* 17(4): 333-344 (2002)

Moree, M. and Ewart, S.: Policy challenges in malaria vaccine introduction. *American Journal of Tropical. Medicine and Hygiene*, 71(Suppl. 2): 2458-252 (2004).

Mosanya, M. E.: The need to involve indigenous health care providers and community– based organization in the control of malaria in the community. *Malaria in Africa*, **1:** 34-50 (2000).

Murphy, S. C. and Breman, J. P.: Gaps in childhood malaria burden in africa: cerebral malaria, neurological sequelae, anaemia, respiratory distress, hypoglycaemia and complications of pregnancy. *American Journal of Tropical Medicine and Hygiene*, 64(1,2 S): 57- 67 (2001).

National Population Commission (NPC): *Nigeria Demographic and Health Survey 2003.* Calverton, Maryland: National Population Commission (2004).

Nuwaha F.: The Challenge of Chroquine-resistant Malaria in sub-Saharan Africa. *Health Policy and Planning*, 16(1): 1-12 (2001).

Pedan, A. (2001). Analysis of count data using the SAS System SUGI pp 247-26

Smith DL, Dushoff J, Snow RW, Hay SI. *The entomological inoculation rate and Plasmodium falciparum infection in African children. Nature 2005;438:492-5.*

Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. *The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature 2005;434:214-7.*

Wellcome News.: *Research Directions in Malaria*. Supplement 6. London (2002).

White, N. J.: Antimalarial drug resistance. *J. Clin. Invest.*, 113:1084-1092 (2004).

WHO.: *The African Malaria Report 2003*. Geneva. WHO/ CDS/MAL (2003).

Woolhouse MEJ, Dye C, Etard JF, et al. *Heterogeneities in the transmission of infectious agents: implications for the design of control programs. Proc Natl Acad Sci USA 1997;94:338-42.*

World Health Organization (WHO): World Malaria Situation, 1994 *Weekly Epidemiological Record*, 72(36): 269-276 (1997).

List of Tables and Figures

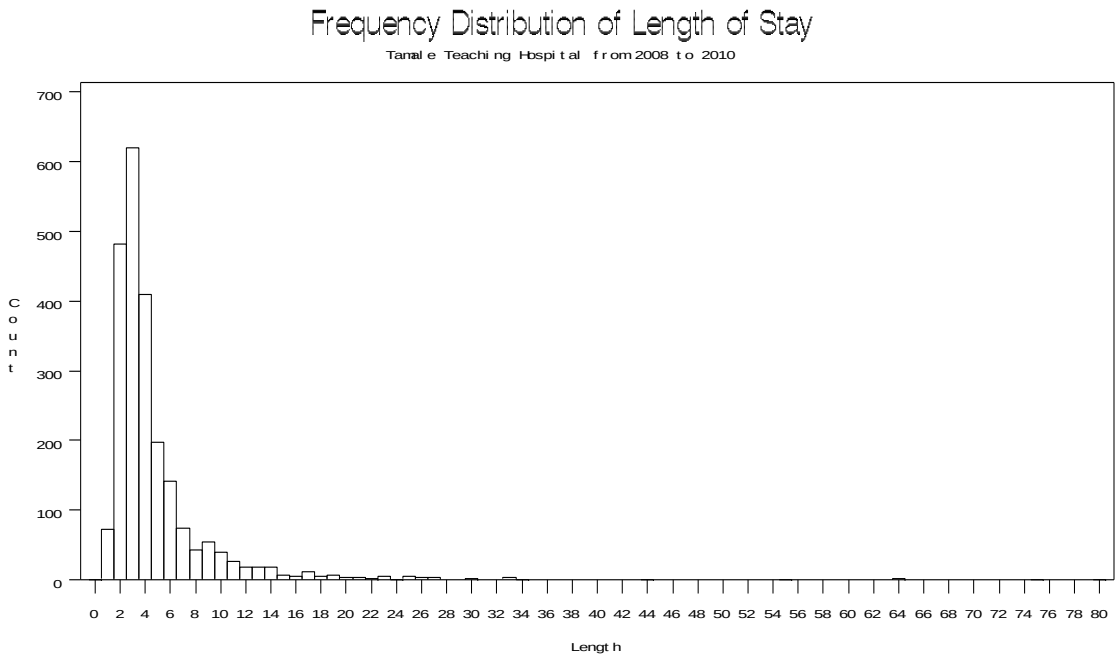

Table 1: Tests for Normality

Test	--Statistic---	-----p Value------		
Kolmogorov-Smirnov	D	0.540794	Pr > D	<0.0100
Cramer-von Mises	W-Sq	173.8762	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	829.9621	Pr > A-Sq	<0.0050

Table 2: General Negative Binomial

Criterion	Reduced[b]	Poisson	Negative Binomial	p=1	p=2[a]
Deviance	6490.15	6198.24	2082.79		
Pearson Chi-Square	11786.0	10692.5	4275.52		
Log-likelihood	6215.65	6362.60	7470.74		
AIC	13858.7	13584.8	11370.5	11411	11371
BIC	13864.4	13647.9	11439.3	11479	11439

[a]The *Negative Binomial* is the same as *General Negative Binomial with p=2,* i.e., $\mu_i + \alpha\mu_i^2$. The coefficient α is estimated by Maximum Likelihood Method (SAS, 2007). The coefficient $\alpha = \frac{1}{r}$ (in eq. 7) and represents the amount of heterogeneity present in the Poisson count data (Hilbe, 1994). The degrees-of-freedom is $n - p$ = 2279 for all models where n = total number of observations and p is the number of parameters in the saturated model.
[b]Model includes the intercept only

Table 3: Model comparison relative to the restricted model

Model	Full-Likelihood	Change in Likelihood	P-value Chi-Square[1]
Intercept only	-6928.34	-	
Poisson	-6781.38	147.00	0.000
Negative Binomial	-5673.25	1255.09	0.000

[1]The 95% critical value of the chi-square distribution with 11 degrees of freedom is 19.6751

Table 4 Analysis of Parameter Estimates from different models fits

(p=2)		Corrected Poisson Model			Poisson Model			Negative Binomial Model		
		Estimate	Standard Error	Pr > Chi sq.	Estimate	Standard Error	Pr > Chi sq.	Estimate	Standard Error	Pr > Chi sq.
Intercept		0.7982	0.1415	<.0001	0.7982	0.0858	<.0001	0.8710	0.1184	<.0001
Outcome		0.7933	0.1399	<.0001	0.7933	0.0848	<.0001	0.7624	0.1280	<.0001
Outcome	0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
ReferralStatus	1	0.6640	0.1756	0.0002	0.6640	0.1065	<.0001	0.5969	0.1550	0.0001
ReferralStatus	0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Age	1	0.0247	0.0048	<.0001	0.0247	0.0029	<.0001	0.0237	0.0047	<.0001
Sex		0.3509	0.1264	0.0055	0.3509	0.0766	<.0001	0.2672	0.1243	0.0316
Sex	0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Season	1	0.0602	0.0411	0.1429	0.0602	0.0249	0.0157	0.0474	0.0384	0.2172
Season	0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Distance	1	-0.1795	0.0398	<.0001	-0.1779	0.0391	<.0001	-0.1779	0.0391	<.0001
Distance	0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Treatment	1	0.1674	0.0481	0.0005	0.1674	0.0292	<.0001	0.1609	0.0462	0.0005
Treatment	0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Outcome*ReferralStat	1	-0.6965	0.1785	<.0001	-0.6965	0.1082	<.0001	-0.6635	0.1659	<.0001
Outcome*ReferralStat	0 0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Outcome*ReferralStat	0 1	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Outcome*ReferralStat	1 0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Outcome*ReferralStat	1 1	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
ReferralStatus*Sex	0 0	-0.3891	0.1306	0.0029	-0.3891	0.0792	<.0001	-0.3018	0.1282	0.0186
ReferralStatus*Sex	0 1	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
ReferralStatus*Sex	1 0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
ReferralStatus*Sex	1 1	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Season*Treatment	0 0	-0.1308	0.0666	0.0496	-0.1308	0.0404	0.0012	-0.1125	0.0637	0.0774
Season*Treatment	0 1	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Season*Treatment	1 0	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Season*Treatment	1 1	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	
Deviance		6200.6454	(1.000)[1]		6200.6454	(2.712)[2]		2083.6478	(0.913)[2]	
Pearson Chi-Square		10697.3001			10697.3001			4277.4605		
Full-Log Likelihood					-6781.3863			-5673.2517		

[1]The value represents the Scale Deviance, [2]Deviance divided by df (=2279)