fjrs.futa.edu.ng

FUTA Journal of Research in Sciences

ISSN: 2315-8239

FUTA Journal of Research in Sciences, 2014 (1): 102-109

LOGISTIC REGRESSION ANALYSIS OF MALARIA PATIENTS TREATED WITH ARTEMISININ-BASED COMBINATION THERAPY (ACT) BY GENDER AND AGE AT OGUI COMMUNITY, ENUGU NORTH LOCAL GOVERNMENT AREA, ENUGU STATE, NIGERIA.

A.S. Ajiboye[1] and G.O. Awoleye[2]

[1]Department of Statistics, The Federal University of Technology, Akure, Nigeria.
E-mail: asajiboye@futa.edu.ng; asajib@yahoo.com

[2]The Federal Polytechnic, Ede, Nigeria. E-mail: rachaelde1@yahoo.com

ABSTRACT

Due to resistance of malaria parasites to traditional drugs used to treat malaria such as chloroquine, new drugs were developed. Among them was Artemisinin- based Combination Therapy (ACT) introduced in 2005. To achieve reduction in malaria incidence, the new drug must be provided to patients in the population, particularly in the rural communities where treatment facilities were usually in short supply. This work examined the consumption of ACT procured through the Global Fund for Malaria for poor and less privileged countries of the world, in aOgui Community, Enugu North Local Government, Enugu State, Nigeria. Descriptive Statistics, Cross-Tabulation and Logistic Regression methods were used. It was found that male and female sexes benefited almost equally from the treatment. While a total of 609 females reported at the health facilities in the community in the period under consideration, 85.0% of them received treatment. For male patients, 701 reported at the health facilities within the same period, 84.3% received treatment. The age group (06months – 14 years) had the largest number of reported cases of malaria in the community during the period. About 738 reported cases were recorded and 96.7% were treated with ACT. Among the age group 'above 14 years', only 572 cases were recorded in the same period and 69.8% received treatment with ACT. The logistic regression model showed that age has effect on the probability of receiving treatment while sex had no effect. Overall, logistic regression showed a good fit for modeling the odds of receiving treatment with ACT.

Keywords:Gender, Artemisinin-based combination therapy, Health facility, Odds ratio, Malaria, logistic regression.

INTRODUCTION

Malaria is common in all parts of Nigeria, with a seasonal difference most striking in the northern part of the country. Malaria is endemic in most parts of Nigeria, with approximately 97% of the population at risk of infection. Current estimates reveal that 110 million cases of malaria occur annually in Nigeria accounting for 63% of all outpatient clinic visits and 30% of all hospital admissions. The predominant parasite species is *Plasmodium falciparum*, which accounts for about 98% of malaria cases in Nigeria (FMOH, 2001). Around the world, malaria is the most significant parasitic disease of humans and claims the lives of more children worldwide than any other infectious disease. Since 1900, the area of the world exposed to malaria has been halved, yet two billion more people are presently exposed (Yotoko and Elisei, 2006). *Plasmodium falciparum* is a protozoan parasite of the species of plasmodium that causes malaria in humans. It is the most dangerous parasite among the species as it accounts for the highest level of complications and mortality. It is transmitted by the female *Anopheles* mosquito. As of 2006, it accounted for 91% of all 247 million human

malaria infections (98% in Africa), and 90% deaths (WHO, 2008). It is more prevalent in sub-Saharan Africa than in other regions of the world. In most African countries, more than 75% of cases were due to *Plasmodium falciparum*, whereas in most other countries with malaria transmission, other plasmodium species predominate (WHO, 2008). Malaria is transmitted largely by the anopheles mosquito and by blood transfusions (Soloqub *et al.*, 2011).

Malaria is one of the leading causes of disease and death in the world. It is estimated that there are 300 to 500 million new cases every year, with 1.5 to 2.7 million deaths worldwide. The severity of malaria and its relationship to other diseases have heightened the urgency of controlling the problem in African countries. Cerebral malaria is now estimated to be responsible for a fatality rate of more than 20 percent of malaria cases, even in urban areas (WHO, 2008).Malaria in sub-Saharan Africa is a problem of dimensions unlike those seen anywhere else in the world today (Duval *et al*., 2010). The magnitude of malaria in Africa is affected by a variety of factors, none of which addressed alone is likely to effect a resolution. It is further compounded by the generally poor social and economic conditions in sub-Saharan Africa. Approximately 80 percent of malaria cases and 90 to 95 percent of malaria-related deaths in the world are estimated to be in Africa. In some areas of sub-Saharan Africa people receive 200 to 300 infective bites per year (Baer *et al*., 2007). At least 300 to 500 million (malaria episodes are treated annually in sub-Saharan Africa. The vast majority of malaria deaths occur there, where malaria also presents major obstacles to social and economic development. Malaria has been estimated to cost Africa more than US$ 12 billion every year in lost GDP, even though it could be controlled for a fraction of that sum (WHO, 2008).

There are several reasons why Africa bears an overwhelming proportion of the malaria burden. This region is also home to the most efficient, and therefore deadly, species of the mosquitoes which transmit the disease. Moreover, many countries in Africa lacked the infrastructures and resources necessary to mount sustainable campaigns against malaria and as a result few benefited from historical efforts to eradicate malaria (WHO, 2003).One of the greatest challenges facing Africa in the fight against malaria is drug resistance. Resistance to chloroquine, the cheapest and most widely used anti-malarial, is common throughout Africa (particularly in southern and eastern parts of the continent). Resistance to sulfadoxine-pyrimethamine (SP), often seen as the first and least expensive alternative to chloroquine, is also increasing in east and southern Africa. As a result of these trends, many countries have to change their treatment policies and use drugs which are more expensive, including combinations of drugs, which it is hoped will slow the development of resistance (FMOH, 2001).The medicine efficacy trials carried out in 2002 and 2004 necessitated policy shift from monotherapy to more efficacious Artemisinin based Combination Therapy (ACT) and this was introduced in 2005 with Artemether- Lumefantrine (AL) as first line treatment for uncomplicated malaria, and artesunate+Amodiaquine (co-packaged) as alternative.

Providing ACTs, Rapid Diagnosis Tests (RDTs) and technical supportive supervision for community case management is crucial to reduce malaria related morbidity and mortality among the vulnerable population segments. Currently, treatment of malaria (fever) is presumptive due to low availability of RDT and microscopy at public health facilities, including client attitude to test before treatment, and service providers to client ratio. These factors influence uptake of ACT for fever cases without clinical diagnosis.Considerable efforts have been undertaken in recent years to increase the access to treatment at community level.

This paper aims to examine proportion of women and men benefiting from the free ACT, examine quantity of ACT consumed by age categories in the community and to fit a logistic regression model to determine if age or gender or both have effect on treatment administration. Because of existing difficulties in accessing treatment, particularly at rural community level, being able to determine the proportion of patients receiving treatment at this level will

enhance efforts of stakeholders to make treatment available to the patients. One of the objectives of the paper is to use logistic regression to achieve this as the model classifies patients into those that received treatment and those that did not. The paper will also determine, using logistic regression, the probability and the odds that a patient will be treated at the rural community level.

THE LOGISTIC REGRESSION METHOD

Logistic regression was developed in the late 1960's and early 1970s as alternative to linear discriminant function analysis and the ordinary least squares regression which were found to be inappropriate for handling dichotomous outcomes due to their strict statistical assumptions (Chao-Ying *et. al*., 2002). Its area of application includes medicine, social sciences, education, market research and so on.

BINARY LOGISTIC REGRESSION

In binary logistic regression, the dependent variable is classified into two, indicating occurrence and non-occurrence of an event. The independent or predictor variable can be categorical or continuous. It is a logit transformation of the dependent variable.

The general regression model is given by

$E(Y/x) = \beta_0 + \beta_1 x$ (1)

Where Y denotes the outcome variable and x denotes the value of the independent variable. E(Y/x) can take on any value as x ranges from $-\infty$ to ∞.

For a dichotomous outcome variable, E(Y/x) will range from 0 to 1. The plot is 'S' shaped, which resembles the plot of a cumulative distribution of a random variable.

Logistic distribution is commonly used in the analysis of dichotomous outcome variables, though some well-known cumulative distributions have been used to approximate the model for E(Y/x) (Cox and Snell, 1989). The logistic model predicts the logit of Y from X. The logit is the natural logarithm of odds of Y, and odds are ratios of probabilities,π ,of Y happening to probabilities (1-π) of Y not happening (Chao-Ying *et al*., 2002). The probability of Y happening is denoted by E(Y/x) and it is equal to π(x). The logistic model can be written as:

$$logit(Y) = ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x \quad (2)$$

Which is the log transformation of the odds.

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \quad (3)$$

The natural logarithm transformation of the odds is necessary to make the relationship between a categorical outcome variable and its predictors linear (Chao-Ying *et. al*., 2002).

The null hypothesis is H_0: $\beta=0$. This states that there is no linear relationship between the outcome variable and the predictor variables. If the null hypothesis is rejected then a linear relationship exists.

Linear regression assumes that ε, the error term, is normally distributed with mean zero and constant variance. This implies that the outcome variable Y has a normal distribution with mean E(Y/x) and constant variance. With dichotomous outcome variable, the situation is different. The outcome variable can be expressed as

$$Y = \pi(x) + \varepsilon \quad (4)$$

ε may assume one of two values. If y = 1, ε = 1 – π(x) with probability π(x). If y = 0, ε = - π(x) with probability (1 - π(x)). With E(ε) = 0 and v(ε) = π(x)(1- π(x)), we can conclude that the conditional distribution of the outcome variable follows a binomial distribution.

DICHOTOMOUS INDEPENDENT VARIABLE

Dichotomous independent variable has covariate values x = 1 or x = 0. The effect of covariate X on the logit of Y can be determined by calculating the difference in the logit function for values x = 1 and x = 0. Using (2), this gives β. The odds of the event happening among individuals with x = 1 is π(1)/(1- π(1)). Similarly, the odds of the event happening among individuals with x = 0 is π(0)/(1 - π(0)). The odds ratio is defined as the ratio of the odds for x = 1 to the odds for x = 0 (Hosmer and Lemeshow, 2002), and is given by

$$OR = \frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))} \quad (5)$$

where OR represents odds ratio. Using equation 3, $OR = e^{\beta}$.

The odds ratio is a measure of association which approximates how much more likely (or unlikely) it is for the outcome to be present among those with x = 1 than among those with x = 0 (Hosmer and Lemeshow, 2002).

The odds ratio approximates the relative risk, $\pi(1)/\pi(0)$. If $\pi(x)$ is small for x = 1 and x = 0, $(1 - \pi(0))/(1 - \pi(1)) \approx 1$. Therefore $OR \approx \pi(1)/\pi(0)$ = Relative risk.

The parameters β_0 and β_1 are estimated by the method of maximum likelihood. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the estimated values of the parameters. Then, $\widehat{OR} = e^{\beta}$.

The odds ratio is usually the parameter of interest in logistic regression. Inferences about the odds ratio are based on the distribution of $ln(\widehat{OR}) = \hat{\beta}_1$, which tends to follow a normal distribution (Hosmer and Lemeshow, 2002). A 100(1-α)% confidence interval for the odds ratio is given by

$$e^{\left(\hat{\beta}_1 \pm z_{(1-\alpha/2)} se(\hat{\beta}_1)\right)} \quad (6)$$

ANALYSIS AND RESULT

The data for this research work was obtained from administrative records of hospitals and health facilities in Ogui Community, Enugu North Local Government, Enugu State, Nigeria. The research is directed towards examining the gender and age composition of Artemisinin-Based Combination Therapy (ACT) consumers in the community.Since monthly health facilities summary report submitted to state's Roll Back Malaria office does not contain information on client's sex and age, health facilities in the community were visited to collect the data. Four months Health Facility data from records (daily diagnosis, preventive and treatment services provision register) were retrieved.

A total of 1311 patients visited the health facilities during the four month period under consideration out of which 738 were children and young people between the age 06months – 14years and 572 were above 14 years. The number of females was 609 while the number of males was 701. Of the 738 children and young people aged 06months-14years, 714 received treatment during the period making a total of 96.7% while only 24 did not receive treatment which is 3.3%. In the case of people above 14 years, 399 received treatment making a total of 69.8% while 173 did not receive treatment which is 30.2%. In the case of gender distribution, 522 of the 609 female patients received treatment and 87 did not while 591 of the 701 male patients received treatment and 110 did not. The percentage treatment among female patients was 85.7% while the percentage among male patients was 84.3%. The percentage that did not receive treatment among female patients was 14.3% while the percentage that did not receive treatment among male patients was 15.7%.

LOGISTIC REGRESSION ANALYSIS

The logistic regression analysis used treatment as the outcome variable and age and sex as independent (covariate) variables. The dependent variable'Treatment' is coded as follows: 0 = 'no treatment given' and 1 = 'treatment given'. The regression will model the odds of receiving treatment in the community. The explanatory variables used in this study were 'age of patients' and 'sex of patients'. Age is classified as follows: 0 = '06 months to 14 years' and 1 = 'Above 14 years'. Similarly for sex, 0 = 'Female' and 1 = 'Male'. The SPSS software package is used to analyze this data. Table 1 shows the SPSS output. From the output, the probability of treatment can be written as

$$P(y=1) = \frac{e^z}{1-e^z} = \frac{1}{1+e^{-z}} \quad (7)$$

where

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 = 3.297 + .227sex - 2.570age \quad (8)$$

Using a .05 level of significance, age and the constant term are found to be highly significant while sex is not.

The sign of the estimated coefficients in the equation above gives an explanation concerning the explanatory variable. Sex is positive which means the sex coded 1 increases the probability of receiving treatment, in this case, 'male'. Age is negative which means the age coded 1 decreases the probability of receiving treatment, in this case, 'above 14 years'. We can say that being male will increase the probability of

receiving treatment and being above 14 years of age will decrease the chance of receiving treatment in this community.

MODEL ASSESSMENT

Intercept only model

The intercept model only is In(odds) = 1.732 as shown in table 2. This model is highly significant at 5% level of significance, which means the intercept can be included in the model. The intercept only model will predict the odds of treatment as Exp(B) = 5.650.

Table 1: Outputs from the SPSS analysis

COVARIATES	BETA COEFFICIENTS	S. E.	WALD	SIG. PROB.	EXP(B)
SEX	.227	.168	1.831	.176	1.255
AGE	-2.570	.227	128.148	.000	.077
CONSTANT	3.297	.218	228.926	.000	27.036

Table 2: Intercept only model

		B	S.E.	Wald	df	Sig.	Exp(B)
	Constant	1.732	.077	501.869	1	.000	5.650

Table 3: model Summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
910.999[a]	.140	.246

Table 4: Variables in the model

		B	S.E.	Wald	df	Sig.	Exp(B)
	Sex(1)	.227	.168	1.831	1	.176	1.255
	Age	-2.570	.227	128.148	1	.000	.077
	Constant	3.297	.218	228.926	1	.000	27.036

Table 5: Model without the covariate Sex

		B	S.E.	Wald	df	Sig.	Exp(B)
	Age	-2.557	.227	127.333	1	.000	.078
	Constant	3.393	.208	267.287	1	.000	29.750

Table 6: Change in -2LogLikelihood When the Covariate Sex is Removed from the Model

Variable		Model Log Likelihood	Change in -2 Log Likelihood	Degrees of Freedom	Sig. of the Change
	Age	-576.438	240.038	1	.000

Table 7: Classification Table

	Observed		Predicted		
			Treatment given?		Percentage Correct
			No	Yes	
	Treatment given?	No	0	197	.0
		Yes	0	1113	100.0
	Overall Percentage				85.0

Full Model

Addition of the two variables: age and sex, gives a chi-square of 198.232 with 2 degrees of freedom and significant probability of 0.00. This tests the null hypothesis of zero beta coefficient, that is, adding age and sex to the model has not significantly improve the model's ability to predict the probability of receiving treatment. The -2loglikelihood statistics is 910.999. This statistics measures how poorly the model predicts the probability of receiving treatment. The smaller the statistics the better the model. Adding the variables has reduced the -2loglikelihood statistics from 1109.23 for intercept only model to 910.999 for the full model. The null hypothesis is rejected and we can conclude that the full model gives a better chance at predicting the treatment probability than the intercept only model.

Sex is contributing only 1.835 to the reduction in the -2loglikelihood statistics with a significant probability of 0.176 which is not significant (table 4). This suggests that we can do without the variable in the model. We use the conditional forward regression method to determine the final model. The results are shown in tables 5,&6.

Table 5 shows the analysis without the covariate Sex. The covariate Age and the constant term are both significant with slight increase in their odds ratios. Table 6 shows the change in -2loglikelihood statistic. This change is highly significant, hence we can conclude that the model without the covariate Sex can be used for the data. This model is given by

$$z = 3.393 - 2.557Age \quad (9)$$

INTERPRETING THE ODDS RATIO.

The odds predicting equation is Exp(Z) = $e^{3.393-2.557x}$. For age (6months – 14years, x = 0), the odds is 29.750, while for age above 14years (x=1), the odds is 2.30712. This means, while age 6months-14years is 29.750 times more likely to receive treatment, age above 14 years is only 2.30712 time more likely to receive treatment.

We can convert odds to probability. For age 6months – 14years,

$$\hat{y} = \frac{odds}{1+odds} = 0.967$$

This means model predicts 96.7% will receive treatment. For above 14years,

$$\hat{y} = 0.6976$$

Model predicts 69.76% will receive treatment.

The odds ratio is given by Exp(-2.557) = .078 (2.30712/29.750). This tells us that the odds of receiving treatment is 0 .078 lower for age above 14years than for age 6months -14years.

CLASSIFICATION

The result of the logistic regression can be used to classify subjects with respect to whether they receive treatment or not. The model predicts the probability of receiving treatment as 0.967 for age '06months – 14 years' and 0.6976 for age 'above 14 years'. This information can be used

to classify subjects. The decision rule is of the form: If the probability of the event is greater than or equal to certain threshold, we predict that the event will take place. SPSS sets the threshold at 0.5. Using this threshold, SPSS classifies subjects into 'Yes Treatment Given' and 'No Treatment Given'. Looking at the classification table, this rule allows us to classify 1113 out of 1113, which is 100% of the subjects where the predicted event is 'Yes Treatment Given' was observed. This is the sensitivity of prediction, the P(correct/event did occur). We can also interpret it as the percentage of occurrences correctly predicted. We also classify the subjects where the predicted event was not observed. The probability is 0 out of 197 which is 0.0%. This probability is called the specificity of prediction and it is P(correct/event did not occur). This is the percentage of non-occurrences correctly predicted. The overall success rate is 85.0%

CONCLUSION

Consumption of ACT was found to be almost evenly distributed among the sexes. The number of female malaria patients was 609 during the period under consideration while the number of male was 701. Out of these numbers, 84% received treatment among the females while 85% received treatment among male patients. However the age group '06months – 14 years'tend to consume more ACT than the age group 'above 14 years'. This age group also recorded larger number of patients at the health facilities during the period, than the other age group. It has a total of 739 patients during the period and represents 56.3% of all malaria patients considered. The age group 'above 14 years', has 572 patients during the period. Of these numbers, 96.7% received treatments for age group '06months – 14 years' while 69.8% received treatment among the age group 'above 14 years'. We can see that while consumption seems to even out among the sexes, patients in the age group '06months – 14 years' are the major consumers of ACT in this community.

The logistic regression model was found to be a good fit for the data. The probability of receiving treatment can be predicted from knowledge of the age group of the patient, but we found that sex of the patient has little or no effect on the probability of receiving treatment. The odds for Age '06months – 14 years' was 29.5, which means patients in this age group are 29.5 times more likely to receive treatment. The odds for age group 'above 14 years' is only 2.31. This means that patients in this age group are only 2.31 times more likely to receive treatment. Converting odds to probability shows that 96.7% of patients in age group '06months – 14 years' will receive treatments while 69.76% of patients in the age group 'above 14 years' will receive treatments.

The odds ratio is 0.078 which shows that the age group 'above 14 years' is .078 times less likely to receive treatment than age group '06months – 14 years'.

The model is also used for classification. With a threshold probability of 0.50, all cases were classified as receiving treatment. This is good for the effort of stakeholders to make ACT available at the rural community level. We can then conclude that the programme of providing free malaria treatment to patients in Ogui Community, Enugu State, Nigeria, is meeting its objective.

REFERENCES

Baer, K., Klot, Z. C., Kappe, S. H., Scnieder, T and Frevert, U (2007). Release of hepatic *Plasmodium yoeli*merozoites into the pulmonary microvasculature "Plos pathogens 3(11), 2171 doi; 10.1371/journal ppat, 0030171.

Chao-Ying, J. P., Lee K. L. and Ingersoll, G.M. (2002). An Introduction to Logistic Regression Analysis and Reporting. The Journal of Education Research 96 (1): 25-31.

Cox, D. R. and Snell, E.J (1989).Analysis of Binary Data, 2nd Edition.Chapman & Hall/CRC.

Duval, L., Fourment, M., Nerrient, E., Rousset, D., Sadeuh, S. A., Goodman, S M., Andriaholinirina, N. V., Randrianarivelojosia, M., Paul, M. R. E, Robert, V., Ayala, F. J. and Ariey, F. (2010). "African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the Laveraniasubqenus" Proceedings National Academic of Science, U.S.A. 107 (2): 10561-6.

FMOH (2001). Federal Ministry of Health National Strategic Plan for roll Back malaria in Nigeria (2001).

Hosmer, D. and Lemeshow, S. (2002). Applied Logistic Regression (Second Edition). New York: John Wiley & Sons, Inc.

Soloqub, L, Kuehn, A., Kern, S., Pryzyborski, J., Schilliq, R and Pradel, G. (2011). Malaria proteases mediate inside-out egress of gametocytes from red blood cells following parasite transmission to the mosquito. Cell Microbial 1462 – 5822.

Szmitko, P. E, Kohn, M. L and Simor, A. E. (2008).“*Plasmodium falciparum* malaria occurring eight years after leaving an endermic area” Diagnostic Microbial Infectious Disease 61 (1): 105 -107.

Tanomsina, N., Imwong, M., Pukrittayakamees, S., Chotivanich, K., Looareesuwan, S., Mayxay, M., Dolecek, C., Hien, T. T., do Rosario, V. E., Arez, A. P., Michon, P., Snounou, G., White, N. J. and Day, N. P. (2007). Genetic analysis of the dihydrofolatereductase - thymidylate synthase gene from geographically diverse Isolates of *Plasmodium malariae*”.Antimicrobial Agents of Chemotherapy 51 (100): 3523 – 30.

WHO (2001). World Health Organisation Regional office for Africa A frame work for Estimating Economic Burden of malaria in the African Region, Harare 2001.

WHO (2003). World Health Organisation/UNICEF Africa malaria report Geneva; 2003 WHO/CDS/MAL/2003; 1093.

WHO (2008). World Health Organisation.World Malaria Report 2008.Switzerland World health organisation, 2008. Pg 99 – 100.

Yotoko, K. S. C. and Elisei, C. (2006). Malaria parasites (Apicomplexa, Haematozoea) and their relationships with their hosts; Is there an evolutionary costs for the specialisation)” Journal of Zoology Systematic Evolution Research 44(4); 265. J. 1439 – 0469, 2006; 00377.