IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 15, Issue 3 Ser. III (May – June 2019), PP 64-70
www.iosrjournals.org

Fitting Ordered Probit Model to Malaria Symptom Dataset

Drinold Mbete[1], Kennedy Nyongesa[2], Joseph Rotich[3]
[1](Department of Mathematics, Masinde Muliro University of Science and Technology, Kenya)
[2](Department, Mathematics, Masinde Muliro University of Science and Technology, Kenya)
[3](Vice Chancellor, Laikipia University, Kenya)

Abstract: *Malaria remains a major infectious disease that affects millions of people. Once infected with Plasmodium parasites, a host can develop a broad range of clinical presentations which result from complex interactions between factors derived from the host, the parasite and the environment. Clinical study of malaria presents a modeling challenge as patients' disease status and progress is partially observed and assessed at discrete clinic visit times. Since patients initiate visits based on symptoms, intense research has focused on identification of reliable prediction for exposure, susceptibility to infection and development of severe malaria complications. Despite detailed literature on malaria infection and transmission, very little has been documented in the existing literature on malaria symptom data collection and fitting the malaria symptom data to a probability distribution and yet these symptoms are common. The symptom data set is then linked to the hidden disease states (state of individual) via the ordered probit model where the Akaike Information Criteria (AIC) is used to estimate the quality of the model and for model selection. The AIC value of 196.2358 confirmed that gamma distribution is best probability distribution to fit malaria symptom data set.*

Keywords: *data, distribution, fit, symptom, symptom dataset, probit*

Date of Submission: 16-06-2019 Date of acceptance: 02-07-2019

I. Introduction

The term Malaria was first used by Dr. Fransisco Torti, but it was not until 1880 that scientists discovered that it was a parasitic disease caused by a unicellular protozoan of the genus Plasmodium which is transmitted by the Anopheles mosquito. There are about 120 species of Plasmodium parasites found in the blood of mammals, reptiles and birds. Four of these species commonly infect humans i.e. *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae.* Of these, *Plasmodium falciparum* mainly found in tropics and sub-tropics is largely responsible for approximately 80% of all malaria cases and approximately 90% of malaria deaths [1]. Malaria still remains a huge public health issue regardless of how many years of research has been con-ducted on how to combat this disease. According to WHO [80], the latest world malaria report released in November 2017 shows that the number of malaria cases reported in the year 2016 was 216 million up from 211 million cases reported in 2015. The report also shows that malaria death estimates in 2016 stood at 445,000 compared to 446,000 deaths in 2015. The high burden of malaria cases in 2016 was in Africa at 90% with 91% cases of deaths reported in children. According to WHO report on malaria cases in Kenya, malaria is one of the leading causes of morbidity and fatality with about 3.5 million children at risk of developing severe malaria, out of which an estimated 34,000 children under five years die every year. The disease is also responsible for 30% of out-patient visits at health centers, economically, it is estimated that 170 million working hours are lost each year because of malaria illness [2].

Malaria is transmitted from person to person by female mosquito of the genus Anopheles. Inside the human host, the parasite undergoes a series of changes. Within half an hour of inoculation of the parasites, the *sporozoites* infect the liver via the blood stream. Here they divide repeatedly into about 30000-40000 *merozoites* over the course of one or two weeks. *Merozoites* are released into the blood stream where they invade red blood cells. Inside these blood cells they grow and divide, eventually causing the rupture of the cell and the release of more *merozoites* which go on to invade new blood cells. A small proportion of merozoites develop into *gametocytes* and can be taken up by a subsequent mosquito bite. Inside the mosquito the parasite undergoes sexual reproduction and then invades the salivary gland. The cycle completes when the infected mosquito bites another human. [3].

Following an infective bite, symptoms appear in about 9-14 days [4]. The initial symptoms of malaria are non-specific and mimic a flu-like syndrome [5]. The classical malaria paroxysm presents three stages i.e. cold stage, hot stage and sweating stage [6]. The cold stage is typically characterized by extreme coldness, shivering and dry pale goose-pimpled skin. During this initial stage that usually lasts between 30-90 minutes, temperature rises gradually to 39^0C [6]. The hot stage start immediately after the shiver has ceased, the stage is

characterized by the patient feeling hot, vomiting, altered consciousness, convulsion and sometimes diarrhea, in this stage temperature rises further to 41^0C [5]. The second stage lasts between 2-6 hours after which the subject enters the sweating stage. In this stage, the patients experience profuse sweating, fall in temperature and sometimes tiredness and sleepy. This stage lasts between 2-3 hours [6]. Without prompt treatment and effective drugs the illness could progress to severe illness often leading to death.

In a study by Martins *et al* [5], there are 19 common symptoms associated with malaria disease which were confirmed and assessed by microscopy, namely; fever, chills, sweating, headache, myalgia, arthralgia, abdominal pain, nausea, vomiting, dizziness, cough, diarrhea, weakness, inappetence, bitter mouth, pallor, coryza, sneezing and score throat. Some of these symptoms are observable symptoms in patients. A healthy individual when he/she is infected with malaria, the disease develops to mild, moderate and final severe depending on the frequency of symptoms he/she has.

II. Literature Review

Several studies have been carried out on malaria for instance Martens *et al* [7] examined the relationship between malaria and environmental and socio-economic variables in Sudan using health production modified model. They used regression analysis method to analyze their results, the regression results showed significant relationships between malaria, rainfall and water bodies while other variables such as Human Development Index, temperature, population density and percent of cultivated areas were not significant while Teklehaimanot *et al* [8] used robust Poisson regression model to model the daily average number of cases in 10 districts of Ethiopia that was associated with rainfall, minimum temperature and maximum temperature as explanatory variable in a polynomial distributed lag model. To improve reliability and generalizability within similar climatic conditions, the districts were grouped into two climatic zones, hot and cold. The results showed that malaria was associated with rainfall and minimum temperature in Ethiopia. In cold districts, rainfall was associated with a delayed increase in malaria cases while the association in the hot districts occurred at relatively shorter lags. The results also showed that in cold districts, minimum temperature was associated with malaria cases with a delayed effect while in hot districts, the effect of mini-mum temperature was non-significant at most lags, and much of its contribution was relatively immediate.

Nkurunziza *et al* [9] modeled the effects of climate on malaria in Burundi using generalized linear models and generalized additive mixed models. The results showed that there was a strong positive association between malaria incidence in a given month and minimum temperature of the previous month. In contrast, the results also showed that rainfall and maximum temperature in a given month have possible negative effect on malaria incidence of the same month. Kakchapati and Ardkaew [10] carried out a study to model the spatial and trends of malaria incidence in Nepal. They used Poisson and negative binomial regression models to fit malaria incidence rates as a function of year and location. Their study showed a steady decreasing trend in malaria incidence, but the numbers of malaria cases were still very high. Sriwattanapongse *et al* [11] used Spearman's correlation between weekly climatic variables (temperatures, relative humidity and rainfall) and malaria to analyze the bivariate relationships between types of malaria parasites and potential climatic factors. A discrete Poisson model was used to identify purely spatial clusters of malaria incidence in the high risk areas. A Poisson regression model combined with distributed lag non-linear model was also used to examine the effects of temperature, relative humidity and rainfall on the number of malaria cases. The residuals were checked to evaluate the adequacy of the model. Sensitivity analysis was performed to ensure that the associations between climate variables and malaria incidences did not change substantially when the degrees of freedom for climate variables were changed while Nkurunziza et al [12] used semi-parametric regression models to model the dependence of malaria cases on spatial determinants and climatic covariates including rainfall, temperature and humidity in Burundi. The results showed that malaria incidence in a given month is strongly associated with minimum temperature of the previous months.

In a study carried out by Kres *et al* [13] to investigate temporal associations' between weekly malaria incidences in 1,993 children < 15 years of age and weekly rainfall. A time series analysis was conducted by using cross-correlation function and autoregressive modeling. The regression model showed that the level of rainfall predicted the malaria incidence after a time lag of 9 weeks (mean = 60 days) and after a time lag between one and two weeks. The analyses provided evidence that high-resolution precipitation data can directly predict malaria incidence in a highly endemic areas while in Kim *et al* [14] estimated the effects of climate factors on *Plasmodium vivax* malaria transmission using generalized linear Poisson models and distributed lag nonlinear models. Their findings showed that malaria transmission in temperate areas was highly dependent on climate factors. Drebel et al [15] carried out a study using logistic regression to estimate and assess malaria prevalence and the use of malaria risk reduction measures and their association with selected background characteristics in South Sudan. The results showed that educational attainment need not be very advanced to affect practices of malaria prevention and treatment. Primary school attendance was a stronger predictor for use of malaria risk reduction measures than any other selected background characteristics. Nath and Mwchahary

[16] analyzed the temporal correlation between malaria incidence and climatic variables using malaria incidence rates in Kokrajhar district of Assam over the period 2001 to 2010. They used linear regressions analysis method to obtain linear relationships between climatic factors and malaria incidence. The results showed that temperature was negatively correlated with non-forest malaria incidence while relative humidity was positively correlated with forest malaria incidence. Wardrop *et al* [17] studied malaria incidence over time and its association with temperature and rainfall in four counties of Yunnan province, China. Seasonal trend decomposition was used to examine secular trends and seasonal patterns in malaria incidence, a Poisson regression with Distributed lag non-linear models was also used to estimate the weather drivers of malaria seasonality. The study showed that there was a declining trend in malaria incidence in all four counties. Bayesian analysis of an epidemiological model of Plasmodium falciparum malaria infection in Ndiop, Senegal is analyzed by Nicole *et al* [18]. The model describes the application of Bayesian calibration of malaria transmission model using longitudinal data gathered from 176 subjects in Ndiop from 1st July, 1993 to 1stAugust, 1994. The model was able to adequately predict Plasmodium falciparum parasitaemia prevalence in the study population that is, during the dry season, the estimated fraction of non-immune subjects went down to 20% and increases upto 80%. The model was also able to predict time-weighted average incidence contributed by non-immune and immune individuals as 0.2 and 0.47 case/day respectively.

In many studies of medical treatment, symptoms are measured repeatedly over time in observation called longitudinal observation. Though we cannot observe directly latent variables, we learn about it by measuring symptom. For the longitudinal models, two latent variables govern disease, one for the probability of experiencing a particular symptom and another for the severity of the experienced symptom. Thus the probability of a symptom and the severity of it depends on both latent variables and observed variables [19]. Latent variables are variables that are not directly observed but are inferred through a mathematical model from other variables that are directly observed or measured. A latent variable model is a statistical model that contains latent i.e. unobserved variables. These variables can either be discrete or continuous. Sometimes latent variables corresponds to aspects of physical reality which could in principle be measured but may not be for practical reason thus in this situation the term hidden variable is commonly used. One advantage of using latent variables is that they can serve to reduce the dimensionality of data. Latent variable link observable data in the real world to symbolic data in the model. Bayesian statistics is often used for inferring latent variables, the common method used inferring latent variables in Bayesian statistics are; Hidden Markov Model (HMM), factor analysis, principal component analysis and Expectation Maximization (EM) algorithm [19].

Zammit *et al* [20] developed an intra-individual consistency model using a logistic-type latent variable model. The latent variable in the model was used to represent the propensity of symptoms and intensity of episodes as these could not be observed directly and needed to be estimated through observation of symptoms episodes in hypoglycemia. The model results showed that there was individual difference in symptom reporting and that adults exhibit distinct intra-individual variability in symptom reporting. Hans et al extended on the model developed by Zammit et al by allowing for different forms of symptom experiencing thresholds between groups variability when symptoms are classified in groups and performing variable selection to determine a predictive model for the effect of patient characteristics and their interactions on symptom consistency. The study was conducted in several health centers in the United Kingdom and data collected from 381 participants aged between 17-75 years. Bayesian estimation was performed for all coefficients in the developed model without grouped symptoms and with grouped symptoms. The analysis shows that a multiplicative form of symptom propensity and episode intensity provides the most suitable symptom experiencing threshold and groups of symptoms show distinct propensity and that gender subjects had significant impact on the consistency of symptom reporting.

Xing et al [21] developed a Bayesian statistical model using latent semi-Markovian state and state-transition statistics for analysis of the time-evolving properties of influenza-like illness with a particular focus on symptoms. Self-reported data from individual student in a college provided daily over a multiple of months was used. The data corresponded to the strength of various infectious-disease-related symptoms reported separately by each individual student. The computation was per-formed using Markov Chain Monte Carlo (MCMC) and statistical analysis per-formed on the daily self-reported symptom scores. The results showed that the weekly pattern (probability of transiting from healthy state to infective state) is typically heightened at either Wednesday or Thursday and tends to be smaller around weekend because of the fact that students are more likely to report symptom during the school week than they are on the weekend.

III. Methodology

3.1 The study area

The research was conducted in Masinde Muliro University of Science and Technology (MMUST) located in Kakamega Town, Kakamega County with an altitude of 1561m above the sea level with a student population of approximately 15000. The levels of malaria risk and transmission intensity in MMUST exhibit

significant spatial and temporal variability related to variations in amount of rainfall, temperature, altitude, topography and human settlement pattern. In this study area, malaria situation is typical of Sub-Saharan Africa making its transmission an all- year -round affair and seasonal variation. The MMUST Health facility records show that between 300-700 cases of malaria are reported each month and this constitutes 75% of all out-patient cases. The main malaria vectors in MMUST are *Anopheles gambiae sensu stricto*, *An. Arabiensis* and *An. Funestus*. *Anopheles gambiae* generally increases in density after the start of the long rains, while *An. funestus* density is seen to vary in direct proportion to the proximity of permanent breeding grounds rather than rainfall [5]. The pick period of malaria incidence occurs from April to August following the main rain season. The malaria cases can either be complicated malaria or un-complicated malaria. For complicated malaria, the following symptoms have been displayed by students; dizziness, hallucination, prostration, loss of consciousness, hyperparasitaemia, pallor, convulsions, low and high blood pressure, coma, convulsions, low and high pulse beat/min, anemia and black quarter fever and dark urine. For uncomplicated malaria, the following non-specific symptoms have been displayed by the students; headache, pains (joint, muscle, abdominal), loose stool, fever, rigors, nausea and vomiting. For confirmatory test of malaria, blood smear (BS) for malaria parasite is carried out [5]. Once a student presents himself/herself to a health officer, the following information are recorded in his/her file; patients complain, history of infection, physical examination for signs and symptoms, impression, investigation of the disease through laboratory test, diagnostic and management of the disease. Depending on the frequency of symptoms a student has, the infection is recorded as mild, moderate and severe for data analysis.

3.2 The Probit Model

Let N = 300 be the number of students who visited the health service with various malaria related symptoms.

Let Z_t be the Hidden state of an individual student at time t. Hidden state is a state in which the individual student is in (disease severity)

Let $O = (x_1, ..., x_p)$ be the number of symptom observed in an individual student. In this study $p = 9$

Let Y_t be the observed and coded symptoms severity of a student at time t i.e. $Y_t \in \{0, ..., M\}^p$ where Yt represents the p symptoms scores reported by a student on day t and M the ordinal scale such that $M = (0, 2, 3, 4)$ with 0 being no symptoms (healthy) and 3 being maximum symptoms.

Table 1: Summary of disease severity Y_t

Observed symptoms	Ordinal scale	Disease severity
0	**0**	**Healthy**
1-3	**1**	**Mild**
4-6	**2**	**Moderate**
6-9	**3**	**Severe**

To link the observed symptoms to the hidden state at time t, we use the ordered Probit model. The ordered Probit model for O_t is derived from a latent and continuous variable Z_t related to a set of explanatory variable according to a standard linear model.

$$Z_t = \varphi O_t + \varepsilon \qquad (1)$$

Where

Z_t is the latent variable and continuous measure of symptom severity is the vector of regression coefficient to be estimated at time t.

φ is the vector of regression coefficient to be estimated

ε is normally distributed with mean zero and unit variance.

O_t is the vector of independent variable describing the symptoms at time t.

Since O_t is drawn conditioned on the hidden state Z_t i.e.

$$O_t \mid Z_t \sim N(\theta_{z_t}, \Sigma_{z_t}) \qquad (2)$$

Let τ be the threshold to be estimated along with the parameter vector φ, then observed and coded symptom severity variable $Y_{(t)}$ is determined as

$$Y_t = \begin{cases} 0 & \text{if } -\infty < Z_{(t)} \leq \tau_1 \\ 1 & \text{if } \tau_1 < Z_{(t)} \leq \tau_2 \\ 2 & \text{if } \tau_2 < Z_t \leq \tau_3 \\ 3 & \text{if } \tau_3 < Z_t \leq \infty \end{cases}$$

Let O_r be the r^{th} component of symptoms of at time t, Y_r be the r^{th} component of observed and coded symptom severity at time t. Since we cannot observe Z_t, instead we observe the categories of Y_t according to the rule

$$Y_r = \begin{cases} 1 & \text{if } O_r > 0 \\ 0 & \text{if } O_r < 0 \end{cases}$$

Assuming that the hidden variable Z_t is linear i.e.

$$Z_t = \tau + \varphi O_t + \varepsilon \tag{3}$$

Denoting the threshold by $\tau_0 < \tau_1 < \tau_2 < \ldots < \tau_{g-2} < \tau_{g-1}$ and the resulting response by Y_r, we observe

$$Y_r = \begin{cases} 0 & \text{if } O_r < 0 \\ 1 & \text{if } 0 < O_r < \tau_1 \\ 2 & \text{if } \tau_1 < O_r < \tau_2 \\ . & . \\ . & . \\ . & . \\ g-1 & \text{if } \tau_{g-2} < O_r < \tau_{g-1} \\ g & \text{if } \tau_{g-1} < O_r \end{cases}$$

where $\tau's$ represent the threshold to be estimated and where each $\tau_g \in \mathbb{R}$ and $\tau_{g-1} < \tau_g$ and $\tau_{g-1} = -\infty$, $\tau_0 = 0$ and $\tau_g = \infty$ therefore

$$Y_r = g \;\; if \;\; \tau_{g-1} < O_r < \tau_g \tag{4}$$

Equation (4) is used to determine the cumulative density function (cdf) as follows;

$$\begin{aligned} P(Y_r \leq p) &= P(Z_t \leq \tau) \\ &= P(\tau + \psi O_p + \epsilon \leq \tau) \\ &= P(\epsilon \leq \tau - \tau - \psi O_p) \quad but \;\; O_p \sim N(0, \Sigma) \\ &= \Phi(\psi O_t). \end{aligned}$$

Where $\Phi(.)$ is the cdf and if we assume cumulative normal, then we obtain the probit model

IV. Results and Discussion

Fitting distribution to data is a procedure of selecting a statistical distribution that best fits to a data set generated by some random process. It is a task of finding a mathematical function which represents in a statistical variable. According to Parzen [22], a statistical data modeling is a field of statistical reasoning that seeks to fit probability distribution to data without knowing what the true model is. One therefore needs to learn the model by a process called statistical model identification which requires judgment and expertise and generally needs an iterative process of distribution choice, parameter estimation and quality of t assessment. R software was used in analysis.

Table 2: Summary statistics for symptom data

Min	Skewness	Mean	Median	SD	1st quartile	3rd quartile	Kurtosis	Max
0.00	-0.65	2.09	2.00	0.97	0.75	2.25	2.20	3.00

The results of Table 1 shows that the average number of symptom data set (Ordinal scales data) is 2.09 with a median of 2.00. It is also evident from the results that the minimum value is 0 and the maximum value is

3. The results also shows that there is a deviation of 0.97 with the lower value of the data being 0.75 and the upper value of the data being 2.25. It is also evident from the results that the data is slightly skewed to the left because the computed value is negative or close to zero which implies that the data is platykurtic because of the Kurtosis value of 2.20 which is less than 3. The kurtosis value quantifies the weight of the tails in comparison to the normal distribution for which the kurtosis equals 3.

Before we fit the dataset into ordered probit model, we divide the symptom dataset into training data (80%) and test data (20%) for analysis then fit the dataset into model so as to produce the deviance statistic to assess the model fit. We train and test the data as shown below;

head(traindata)

	Status.of.student	*Number.of.symptoms*	*Age*	*Ordinal.scale*
1	*moderate*	5	20	2
2	*moderate*	6	21	2
5	*severe*	11	21	3
7	*moderate*	6	23	2
8	*severe*	9	24	3
9	*mild*	3	22	1

From the train data, we can clearly see that some values of symptom dataset have been left out for testing the dataset.

head(testdata)

	Status.of.student	*Number.of.symptoms*	*Age*	*Ordinal.scale*
3	*mild*	3	23	1
4	*severe*	9	22	3
6	*mild*	2	21	2
23	*mild*	1	18	1
32	*moderate*	6	24	2
40	*severe*	11	22	3

The test dataset has values that were not captured in the train dataset. We use the test dataset to validate the model by predicting the values in the test dataset.

In R software;

prediction

$pred < -predict(results, testdata)$

head(pred)

Dataset	3	4	6	23	32	40
Ordinalscale	1	3	2	1	2	3

The results shows that the values predicted are the same as the test data value, therefore the model is valid and can be fitted. We now fit the ordered probit model using the train dataset using Stata.

```
. oprobit ordinalscale age    numberofsymptoms

Iteration 0:   log likelihood = -325.71232
Iteration 1:   log likelihood = -123.89726
Iteration 2:   log likelihood = -110.17843
Iteration 3:   log likelihood = -109.30572
Iteration 4:   log likelihood = -109.30205
Iteration 5:   log likelihood = -109.30204

Ordered probit regression                         Number of obs   =        300
                                                  LR chi2(2)      =     432.82
                                                  Prob > chi2     =     0.0000
Log likelihood = -109.30204                       Pseudo R2       =     0.6644

------------------------------------------------------------------------------
ordinalscale |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   -.026457   .0436542    -0.61   0.544    -.1120178    .0591037
numberofsy~s |   .9294152   .0634755    14.64   0.000     .8050055    1.053825
-------------+----------------------------------------------------------------
       /cut1 |   -.732962   1.001887                     -2.696625    1.230701
       /cut2 |   3.299554    .957663                      1.422569    5.176539
       /cut3 |   5.724882   1.002045                      3.760909    7.688855
------------------------------------------------------------------------------
```

The results shows that both number of symptoms and age are statistically significant. The value of R^2= 0.6644 implies that 66/% of the variation in the dependent variable has been explained by the independent variable. The ordered probit regression coefficient gives the change in the z-score or probit index for one unit change in the predictor. For instance, a one unit increase in the number of symptoms, the z-score increases by 14.64 and for each one unit decrease in the age of the student, the z-score decreases by -0.61. The results also shows that the chi-squared test statistic of 432.82 with 2 degrees of freedom is associated with a p-value of less than 0.001 indicating that the overall effect of rank is statistically significant and model fits significantly better than an empty model.

Distri-bution	Para-meter	Estimate	Statement	AIC	BIC	Leglikelihood
Gamma	Shape	7.4318	1.8033	196.2358	197.2056	−96.1188
	Rate	0.0036	0.0008			
Weibull	Shape	1.3622	0.3307	197.8707	198.8405	−96.9354
	Scale	1369.2862	304.5457			
Lognormal	Meanlog	6.7673	0.2846	200.1029	201.0727	−98.0514
	sdlog	0.9858	0.2012			

The Akaike information criteria (AIC) was also used to estimates the quality of the model and as a measure of the relative quality of statistical models for a given set of data. By using AIC values of results, Gamma distribution had the lowest AIC value of 196.2358 followed by the Weibull distribution with a value of 197.8707 and then the Lognormal distribution with a value of 200.1029. This confirmed that Gamma distribution to be the best fitting distribution for the symptom dataset.

V. Conclusion

In this study, we presented methods for developing the model and the fitting of the ordered probit model to symptom data in order to determine the appropriate probability distribution for symptom data set The Gamma, Weibull and Lognormal distributions were selected since the data was positively skewed. The gamma distribution was identified as the best fit for the model because of its small value of AIC.

References

[1]. Musa MI, SHohaimi S, Hashim NR, Krishnarajah, (2012). A climate based distribution model of malaria transmission in Sudan. Geospatial Health 7(1), 27-36.

[2]. Wang Xiang, David Sontag, and Fei Wang, (2014). Unsupervised learning of disease progression Models. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge Discovery and data mining, pages 85-94.

[3]. Ceusters W and Smith B (2009). Malaria Diagnosis and the Plasmodium Life Cycle: the BFO Perspective. Nature Precedings.

[4]. Gilles H.M and D.A Warrell (1993). Bruce-Chwatts Essential Malariology, Third Edition. Edward Arnold, London.

[5]. Martins et al.(2015). Clustering symptoms of non-severe malaria in semi-immune Amazonian patients. PeerJ

[6]. Alessandro Bartoloni and Lorenzo Zammarchi (2012). Clinical Aspect of Un- complicated and Severe Malaria, Mediterranean Journal of Hematology and Infectious diseases.

[7]. Maruotti A. (2013). Modelling longitudinal data with non-ignorable dropouts using hidden Markov models: A conditional maximum likelihood approach. Under revision.

[8]. Li H-M FL-Y, Wang P and Yan J-Z.(2013). Hidden Markvo Models Based Research on Lung Cancer Progress Modeling. Research Journal of Applied Sciences, Engineering and Technology. 6(13):2470-3

[9]. A.Ngwa and W.S.shu (2000). A mathematical model for endemic malaria with variable human and mosquito populations. Math. Comput. Mathematical modeling, 32:747-763.

[10]. Jones G., W. O. Johnson, W. D. Vink, and N. French. (2012). A framework for the joint modeling of longitudinal diagnostic outcome data and latent infection status: Application to investigating the temporal relationship between infection and disease. Biometrics, 68(2):371-379, June 2012.

[11]. Lee H-K, Lee J, Kim H, Ha J-Y and Lee KJ.(2013). Snoring detection using apiezo snoring sensor based on hidden Markov models. Physiological measurement. 34(5): 41-45.

[12]. Nicole Cancre (2000). Bayesian analysis of an epidemiological model of Plasmodium falciparum malaria infection in Ndiop, Senegal. American Journal of Epidemiology, Oxford University Press, 152(8),pp 760-770.

[13]. Koella JC (1991). On the use of mathematical models of malaria transmission, Acta Trop. 1991 Apr; 49(1):1-25.

[14]. Kermack W.O and A. G. McKendrick,(1927). A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. A 115, 700-721.

[15]. Dondorp AM and Day NP (2007). The treatment of severe malaria. Trans R Soc Trop Med Hyg, 2007, vol. 101 (pg. 633-634)

[16]. Nadjm B, Behrens RH (2012). "Malaria: An update for physicians". Infectious Disease Clinics of North America 26 (2): 24359.

[17]. Tumwine J., Mugisha J.Y.T., and Luboobi L.S., (2008). A mathematical model for the dynamics of malaria in a human host and mosquito vector with temporary immunity, Applied Mathematics and Computation, 1953-1965.

[18]. Nath, D.C. and Mwchahary, D.D. (2013) Association between Climatic Variables and Malaria Incidence: A Study in Kokrajhar District of Assam, India: Climatic Variables and Malaria Incidence in Kokrajhar District. Global Journal of Health Science, 5, 90.

[19]. WHO (2012).World Malaria Report.

[20]. Nkurunziza, H., Gebhardt, A. and Pilz, J. (2011) Geo-Additive Modeling of Malaria in Burundi. Malaria Journal, 10, 234.