A Stochastic Model of Some Endemic Infections: Case Study Based on the Medical Records of Gbagada General Hospital, Lagos State, Nigeria.

E. B. Nkemnole * *and R. O. Osunkeye* *

Department of Mathematics, University of Lagos, Nigeria

Abstract: This research investigates the endemic level and (knowledge of) persistence time of endemic diseases in a population using a Stochastic Model such as Markov process which is a continuous time and discrete state space that requires the Monte Carlo Simulation for the desired results to be gotten. It assesses some areas of active research in efficient procedures for simulation in health sector and addresses the influence of gender as regards to the average days a particular disease dies out. There is also an emphasis on the average population to be infected on a monthly basis. The data used for this study is obtained from the medical records that runs from January 2012 through December 2012 of General Hospital Gbagada, Lagos State which comprises of children from ages 0 to 12years and adults from ages 13years and above. The Monte Carlo simulation carried out shows the trend line and equations of the data. Empirical analysis showed a significant relationship between gender and persistence time of endemic diseases in a population.

Keywords: continuous time Markov chain, epidemic, stochastic epidemic models

1 Introduction

Disease is an abnormal condition that affects the body of an organism. It is often constructed as a medical condition associated with specific symptoms and signs. It may be caused by internal dysfunctions, such as autoimmune disease. In humans, disease is often used more broadly to refer to any condition that cause pain, dysfunction, distress, social problems, or death to the person afflicted or similar problem for those in contact with the person. In this sense, it sometimes includes injuries, disabilities, disorders, syndromes, infections, isolated symptoms, deviant behaviors, and atypical variations of structure and function, while in other contexts and for other purposes these may be considered in distinguishable categories. Diseases usually affect people not only physically, but also emotionally as contracting and living with diseases can alter ones perspective of life and their personality.

World Health Organization (WHO) [1] reported that infectious disease crisis of global proportions is today threatening the ground-breaking accomplishments that have been recorded in world health thereby increasing life expectancy. Infectious diseases are now the worlds biggest killer of children and young adults. They account for more than 13 million deaths a year one in two deaths in developing countries. Over the next hour alone, 1500 people will die from an infectious disease- over half of them are children under five. The rest will mostly be working-age adults.

This report argues that we have a window of opportunity to make dramatic progress against ancient diseases, and to establish an early warning system to protect us from new and unexpected diseases. If this is left undone, the result will be the emergence of increased-drug-resistant bacteria and viruses that will pose a major challenge to science, economics and politics. The spread of infectious disease through a population may be modeled as a stochastic process. When infection persists in the population for a long time, the disease is said to be endemic.

Endemic diseases are ones that are always present in a community, usually at a low, more or less constant frequency. Malaria, arthritis, and high blood pressure are examples. Endemic diseases are in most communities around the world (all the time).Cholera is a strictly epidemic disease; it comes and goes but doesn't stay in a community for extended periods.

* Corresponding author e-mail: enkemnole@unilag.edu.ng, raphealosunkeye@rockemail.com

In addition, there is a form Typhus (there are actually four forms, spread by different bugs which bite people) called "endemic typhus". Endemic typhus is carried by fleas. When a flea lands on a human, it may defecate as it feeds. When the person scratches the itchy spot where the flea was feeding, the bacteria-laden faeces are scratched into the skin, thus causing infection. The causative bacteria are called Rickettsia typhus. Endemic typhus occurs most commonly in warm, coastal regions. An infection is said to be endemic in a population when that infection is maintained in the population without the need for external inputs. For example, chickenpox is endemic in Nigeria likewise malaria. For an infection that relies on person-to-person transmission to be endemic, each person who becomes infected with the disease must pass it on to one other person on average. Assuming a completely susceptible population, that means that the basic reproduction number) of the infection must equal 1. In the population with some immune individuals, the basic reproduction number R_0 multiplied by the proportion of susceptible individuals in the population S must equals 1. This takes account probability of each individual to whom the disease may be transmitted actually being susceptible to it, effectively discounting the immune sector the population. For the disease to be in an endemic steady state:

$$R_0 \times S = 1 \tag{1}$$

In this way, the infection neither dies nor does the number of infected people increases exponentially, then the infection is said to be in an endemic steady state. An infection that starts as an epidemic will eventually either die out (with the possibility of it resurging in a theoretically predictable cyclical manner) or reach the endemic steady state, depending on the number of factors, including the virulence of the disease and its mode of transmission .If a disease is in endemic steady state in a population, the relation above allows us to estimate the R_0 of a particular infection.

2 Literature Review

Mathematical modeling of infectious diseases has a long history; see, in particular, [2]. Early modeling contributions for infectious disease spread were often used for specific diseases. The starting point is generally taken to be a paper by [3] on the prevention of smallpox by inoculation; an account of his model-based analysis of data can be found in [4,5] modeled the transmission of malaria. One of the first more general research was conducted by [6]. Later important contributions were made by [7] and [8], both also considering stochastic models.

Early models were often deterministic with questions such as: Is it possible that there is a big outbreak infecting a positive fraction of the community?, How many will get infected if the epidemic takes off?, What are the effects of vaccinating a given community fraction prior to the arrival of the disease?, What is the endemic level? As problems were resolved, the simple models were generalized in several ways towards making them more realistic. Some such extensions were for example to allow for a community where there are different types of individual, allowing for non-uniform mixing between individuals (i.e. infectious individuals dont infect all individuals equally likely), for example due to social or spatial aspects, and to allow seasonal variations.

Another generalization of the initial simple deterministic epidemic model was to study stochastic epidemic models. A stochastic model is of course preferable when studying a small community. But, even with a large community, which deterministic models primarily are aimed for, some additional questions can be raised when considering stochastic epidemic models: What is the probability of a major outbreak?, and for models describing an endemic situation: How long is the disease likely to persist (with or without intervention)? Later stochastic models have also shown to be advantageous when the contact structure in the community contains small complete graphs; households and other local social networks being common examples. Needless to say, both deterministic and stochastic epidemic models have their important roles to play and deterministic and stochastic models are used for epidemiological modeling however, the focus in the present paper is on stochastic model. The stochastic model is a Markov population process with continuous time and discrete state space.

[9b] averred that stochastic models should be established and studied for several endemic infections with demography. Approximations of quasi-stationary distributions and of times to extinction are derived for stochastic versions of Susceptible and Infected (SI), Susceptible Infected Susceptible (SIS), Susceptible Infected Recovered (SIR), and Susceptible Infected Recovered Susceptible (SIRS) models. The approximations are valid for sufficiently large population sizes. Conditions for validity of the approximations are given for each of the models. These are also conditions for validity of the corresponding deterministic model. It is noted that some deterministic models are unacceptable approximations of the stochastic models for a large range of realistic parameter values.

For him, SIS model without demography is an univariate model. The stochastic version of this model is a finite-state birth-death process with an absorbing state at the origin. The goal of the mathematical analysis of this model is to find approximations of the quasi-stationary distribution and of the time to extinction. He used the Kolmogorov equation by introducing four parameters, namely the expected population size N in case where there are no infected individuals, the contact rate β , the death rate per susceptible individual μ, and an additional death rate μ_1 such that the death rate per infected individual is $\mu + \mu_1$. Among these, N is a large positive integer, β and μ are positive, and μ_1 is non-negative.

In addition to the four parameters N, β μ and μ_1 introduced, he used γ_1 to denote the recovery rate per infected individual. The SIS model takes the form of a bivariate Markov population process. The Kolmogorov forward equations for the state probabilities are:

$$P_{si}(t) = P\{S(t) = s, I(t) = i\} \tag{2}$$

This can be written as

$$P'_{si}(t) = \lambda_1(s-1,i)P_{s-1,i}(t) + \mu_1(s+1,i)P_{s+1,i}(t) + \mu_2(s,i+1)P_{s,i+1}(t) + v_1(s-1,i+1)P_{s-1,i+1}(t) + v_2(s-1,i+1)P_{s-1,i+1}(t) - k(s,i)P_{s,i}$$

where,

$$k(s,i) = \lambda_1(s,i) + \mu_1(s,i) + \mu_2(s,i) + v_1(s,i) + v_2(s,i) \tag{3}$$

[9a] also formulated a stochastic SIR model with demography. The model is a Markov population process with three state variables, $S, I, and R$, standing for the number of susceptible, infected, and recovered individuals, respectively. The recovered individuals are assumed permanently immune to additional infections. He used five basic parameters as for the SIS model namely: the expected population size N if all individuals are susceptible, and the death rate per susceptible or removed individual μ, the additional death rate per infected individual μ_1, the contact rate β, and the recovery rate per infected individual γ_1.

The state probabilities are defined by:

$$P_{sir}(t) = P\{S(t) = s, I(t) = I, R(t) = r\} \tag{4}$$

The Kolmogorov forward equation for these probabilities can be written as

$$P'_{sir}(t) = \lambda_1(s-1,i,r)P_{s-1,i,r}(t) + \mu_1(s+1,i,r)P_{s+1,i,r}(t) + \mu_2(s,i+1,r)P_{s,i+1,r}(t)$$
$$+ \mu_3(s,i,r+1)P_{s,i,r+1} + v_2(s+1,i-1)P_{s+1,i-1,r}(t) + v_3(s-1,i+1)P_{s,i+1,r-1}(t) - k(s,i,r)P_{sir}$$

where $k(s,i,r) = \lambda_1(s,i,r) + \mu_1(s,i,r) + \mu_2(s,i,r) + v_2(s,i,r) + v_3(s,i,r)$

2.1 Statistics as Regard Infections

[10] maintained that the reporting interval of infectious diseases is often determined as a time unit in the calendar regardless of the epidemiological characteristics of the disease. No guidelines have been proposed to choose the reporting interval of infectious diseases. His study aimed at translating coarsely reported epidemic data into the reproduction number and clarifying the ideal reporting interval to offer detailed insights into the time course of an epidemic. He derived a corrected expression for this quantity and proposes simple algorithms to estimate the effective reproduction number as a function of time, adjusting the reporting interval to the generation time of a disease and demonstrating a clear relationship among the generation-time distribution, reporting interval and growth rate of an epidemic.

[11] is of the view that the statistical method to determine the reporting interval is density estimation, which may suggest a stochastic model for this project. To interpret the time course of an endemic, case notifications are used to estimate a key variable that characterizes transmissibility with time. The effective reproduction number at time t, R_t, defined as the average defined as the average number of secondary cases per primary case at time t (for $t > 0$), is a useful measure to inform about the transmission potential of a disease and indications of the expected number of secondary transmissions and of control efforts required to curb the epidemic. Although the most precise reporting interval (e.g. reporting in a continuous time scale) would certainly yield the most ideal interpretation of the transmission dynamics, it is often impractical to get data and analyze on an hourly or daily basis.

[12] proffered that a global malaria eradication effort will require massive changes to complex web of interconnected biological systems. The optimal path to eradication is intrinsically unpredictable because of the potential for parasites and vectors to evolve, the waxing and waning of human immunity, and behavioral changes in human and vector populations. The range of conditions that favor malaria transmission are so varied and diverse that decisions and plans cannot be based solely on the evidence that has been acquired in randomized control traits conducted in only few settings. To succeed, eradication will require a strategic plan that is constantly updated with the surveillance, monitoring, and evaluation data. Moreover, planning processes involve some sort of conceptual model, and this model will necessarily consider many potential sources of uncertainty. Rational quantitative mathematical models provide that best way to synthesize information, quantify uncertainty, and extrapolate current knowledge. Such models can provide critical quantitative insights that are not otherwise possible.

The objective of this study is to investigate the endemic level and knowledge of persistence time of endemic diseases in a population by the use of Stochastic Model such as Markov process with continuous time and discrete state space with a view to determining the proportions of population infected and equally know the duration until the disease dies out. The outline of the remainder of this article is organized as follows: Section 3 describes the data and discusses the method of analysis. Results of our analysis are presented in Section 4, while Section 5 concludes.

3 Data and Methodology

3.1 Data

The data used for this study is obtained from the medical record of the General Hospital Gbagada, Lagos State, comprising of children from age 0 to 12years and adults from age 13 years and above. Research was carried out on the following: Diseases of children from the age 0 to 12 years which include: Neonatal jaundice (NNJ), Neonatal sepsis (NNS), Bronchopneumonia (BPN), Plasmodiasis (PD), Gastroenteritis (GN), Malaria; diseases of adults from the age of 13 years and above which include: Critical Cardiac Failure (CCF) Cardiovascular (CV), Chronic Kidney Disease (CKD), Hyperglycemia (HG), Appenditis, Oral sepsis (OS).

3.2 Method

The study utilizes the Markov population process with continuous time and discrete state space. In a continuous time Markov chain (CTMC), time is continuous, but the state variable is discrete. Markov chain can be thought of as a directed graph of states of the system. The difference is that, rather than transitioning to a new (possibly the same) state at each time step, the system will remain in the current state for some random; exponentially distributed, amount of time and then transition to a different state. According to [13], the CTMC epidemic processes are defined on a continuous time scale, $t \in [0, \infty)$, but the states $S(t)$, $I(t)$ and $R(t)$ are discrete random variable i.e

$$S(t), I(t), R(t) \in \{0, 1, 2, ...N\}$$

Here, the stochastic process depends on the collection of discrete random variables and their associated probability functions, $P(t) = (p_0(t), ..., p_N(t))^T$

In our study, attempt was made in formulating the CTMC based on the SIR epidemic models to the Diseases. To numerically compute a sample path of a CTMC model, we used the fact that the intervene time has an exponential distribution. This follows from the Markov property.

The probability distribution function (PDF) of an exponential distribution is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

The cumulative distribution function (CDF) of an exponential distribution is

$$f(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{6}$$

Let $U = F(x)$
Make x the subject of formula,

$$x = \frac{-1}{\lambda} \log(1 - u) \tag{7}$$

U denotes the generated random numbers.
$\lambda = \frac{1}{mean}$

$$Thus X = \frac{-1}{\lambda} \log(RAND(0, 1)) \tag{8}$$

Consequently, the Microsoft Office Excel Package with the Monte Carlo simulation add-in was used to carry out Monte Carlo simulation for the diseases-data thereby generating sample paths and the probability distribution associated with CTMC SIR epidemic models. The output gives the trend line and trend equation, that is; $= 11.439e^{-0.011x}$, histogram and cross tabulation which explains the average, minimum, maximum and standard deviation for both the duration of time the diseases die out and the population to be infected.

4 Study Results

Running the simulation on Monte Carlo add-in package on Microsoft Excel with 1000 repetitions, that is, 1000 sample spaces into the future prior to the original data gotten from the medical record of the General Hospital, Gbagada; the results are presented on Figures 1-12 and Tables 1-2 shown below with its respective summary statistics for Male and Female infected adults and Male and Female infected children.

Table 1 is a presentation of the approximate average days that a disease would die out in a male and female patient respectively: Appendicitis (12, 11); CCF (15, 12); CVD (27, 17); Chronic Kidney Disease (11, 10); Hyperglycemia (12, 12); Oral sepsis (11, 9).

Table 1: Estimates of Duration of Diseases in infected Male and Female adults with corresponding infections

| Statistics	Table 1											
	APPD		CCF		CVD		CKD		HG		OS	
	M	F	M	F	M	F	M	F	M	F	M	F
Average	12.196	10.647	19.816	12.447	26.936	16.622	11.236	10.332	10.704	9.331	12.455	12.204
Std. Dev.	0.4365	1.2944	3.1748	1.4836	12.7936	0.456	1.8731	0.4539	0.2430	5.1005	1.5597	0.4441
Maximum	13.005	12.063	45.449	24.065	114.000	19.632	25.442	14.892	12.791	18.189	26.903	12.682
Minimum	11.442	3.481	12.280	11.117	14.432	16.144	9.585	9.924	10.470	0.024	11.181	9.915
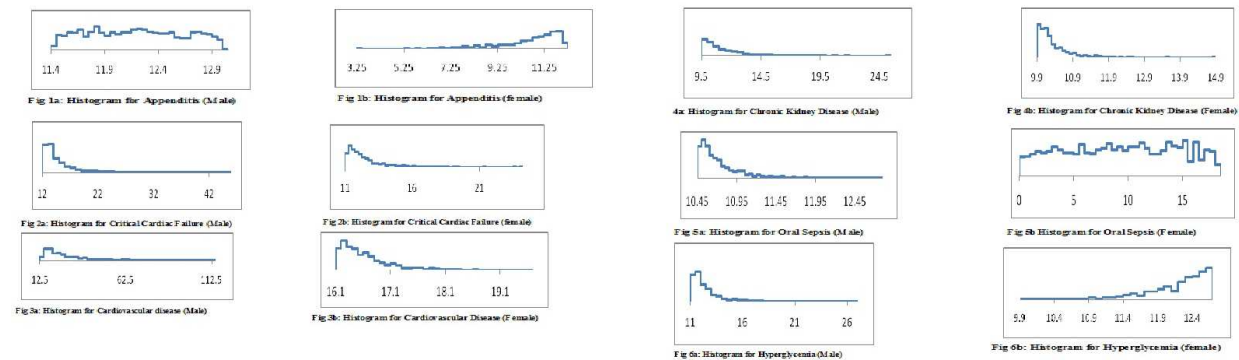
Fig. 1: Graphical representation of Duration of Diseases in infected Male and Female adults with corresponding infections.

Table 2 is a presentation of the approximate average days that a disease would die out in a male and female children respectively: neo natal jaundice (7, 6); neo natal sepsis (5, 7); bronchopneumonia (5, 4); Plasmodiasis (3, 3); Gastroenteritis (4 , 4); Malaria (2, 5).

Table 2: Estimates of Duration of Diseases in infected Male and Female children with corresponding infections

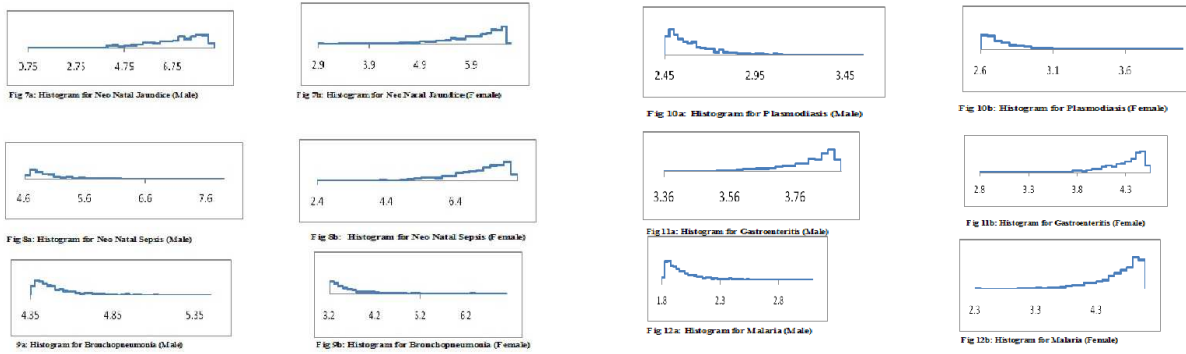| Statistics	Table 2											
	NNJ		NNS		BPN		PLASM		GN		MALARIA	
	M	F	M	F	M	F	M	F	M	F	M	F
Average	6.674	5.929	5.081	6.990	4.508	3.591	2.603	2.755	3.793	4.279	4.664	9.915
Std. Dev.	1.3846	0.5989	0.4443	0.9296	0.1473	0.4296	0.1498	0.1563	0.0916	0.2244	1.899	0.3861
Maximum	8.338	6.601	7.841	8.071	5.441	7.096	3.547	3.964	3.888	4.515	3.081	5.082
Minimum	0.755	2.975	4.656	2.499	4.363	3.221	2.460	2.610	3.361	2.836	1.822	2.321

Fig. 2: Graphical representation of Duration of Diseases in infected Male and Female Children with corresponding infections
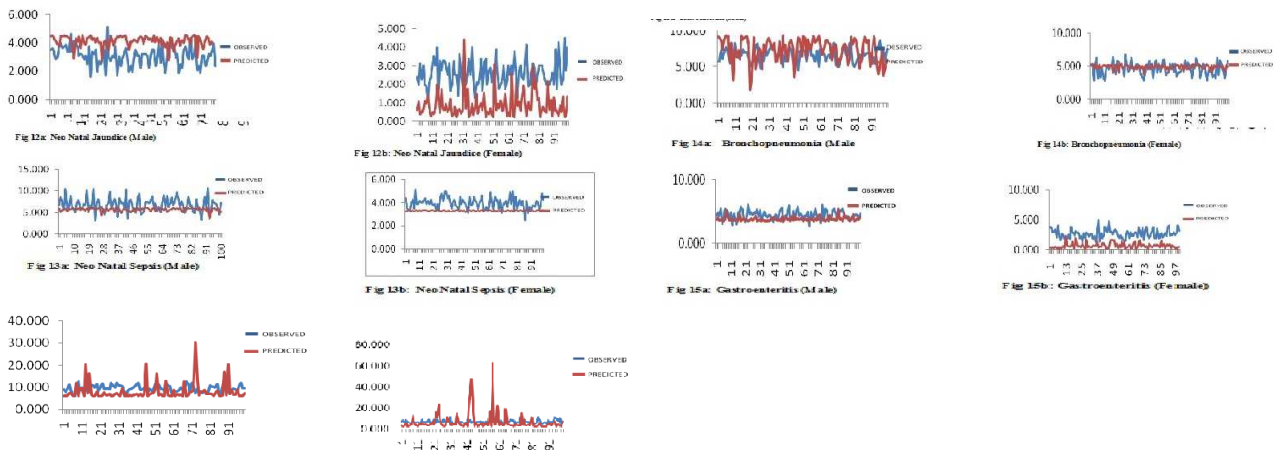

Fig. 3: (a) Histogram for Cardiovascular disease (Male) (b) Histogram for Cardiovascular Disease (Female)

4.1 Forecast

The above procedure is used to forecast the average population to be infected with the diseases on a monthly basis which is categorized into the children category of ages 0 to 12years and adults from ages 13 years and above. The results are presented on Tables 3-4 shown below with its respective summary statistics. The observed and predicted monthly incidence are plotted and displayed in Figures 8-16 (red line represents predicted disease while blue line represents observed disease).

Table 3 is a presentation of the forecast on the average population to be infected on a monthly basis in male and female adults respectively: Appendicitis (1, 1); Critical Cardiac Failure (2, 2); cardiovascular disease (2, 2); Chronic Kidney Disease (2, 1).

Table 3: Estimates of forecast on Male and Female adults to be infected with the diseases on a monthly basis

Table 3								
Statistics	APPD		CCF		CVD		CKD	
	M	F	M	F	M	F	M	F
Average	1.491	0.915	1.509	2.211	2.080	1.671	1.643	1.271
Std. Dev.	0.2238	0.2665	0.0909	0.1604	0.0063	0.0294	0.0787	0.2912
Maximum	3.306	1.410	2.438	3.301	2.087	1.876	1.724	4.523
Minimum	1.294	0.000	1.426	2.060	2.039	1.642	1.248	1.040

5 Conclusion

There has been a huge increase in world's population which equally means there's huge increase in medical supplies at various hospitals. Sadly, in some areas of the world, the medical centers either uses old medical equipment which serves few patients and some equipment are not available to tackle some diseases.

The study is centered on the investigation of endemic level and knowledge of persistence time of endemic diseases in a population by the use of Stochastic Model such as Markov process with continuous time and discrete state space with a view to determining the proportions of population infected and time until the disease dies out. The data from medical records of General Hospital Gbagada are used for the study. We have focused on presenting results for a fairly simple stochastic epidemic model; the reasons being that even in simple models results are far from trivial. The Microsoft Office Excel Package with the Monte Carlo simulation add-in was utilized to carry out Monte Carlo simulation for the diseases-data thereby generating sample paths and the probability distribution associated with the CTMC SIR epidemic model. The output is an indication that there was a significant association between gender and duration in which a disease dies out, and also between diseases and population of male and female to be infected on a monthly basis respectively.

With this research, we recommend that government should endeavor to make available more medical supplies to medical institutes and liaise with developed country in getting specialist in medical fields so as to get the work done as soon as possible, lastly, government should give study grants to medical doctors to get medical trainings because going by the results of the research, we draw out that there is going to be an exponential increase in population to be infected with various diseases compared to what has been experiences in time past.

References

[1] World Health Organization (WHO): THE BIGGEST KILLER OF THE YOUNG, *http://www.who.int/infectious-disease-report/pages/textonly.html*

[2] N.T.J. Bailey,*The Mathematical Theory of Infectious Diseases and its Applications.,* Griffin , London, 1975.

[3] D. J.,Daley and Gani, J.,*Epidemic Modelling An Introduction.* Cambridge Studies in Mathematical Biology, Vol. 15. Cambridge Univ. Press, Cambridge. 1999.

[4] R. Ross, *The Prevention of Malaria, 2nd edition. London*: Murray, 1911.

[5] W. O. Kermack, and McKendrick, A. G. *Contributions to the mathematical theory of epidemics,* part I. Proceedings of the Royal Society of London A,115, 700721, 1927.

[6] M. S.Bartlett, *Some evolutionary stochastic processes.* of the Royal Statistical Society B11, 211229, 1949.

[7] D.G.. Kendall, *Kendall, Deterministic and stochastic epidemics in closed populations. Proc.*

[8] I.Nasell. Endemicity, *persistence, and quasi-stationarity. In: Castillo-Chavez, C. with Blower, S., van den Driessche, P., D. Kirschner, D., Yakubu, A.* In Mathematical Approaches for Emerging and Reemerging Infectious Diseases An Introduction. Springer-Verlag, New York, pp. 199227, 2002a,

[9] I. Nsell, *Stochastic models of some endemic infections.* of Mathematics, The Royal Institute of Technology, 179(1):1-19, 2002b.

[10] I. Nishirura, *Theoretical Epidemiology, university of Utrecht,* The Netherlands; Journal of Royal Society: interface 2010 Feb 6; 7(43):297- 307, 2010.

[11] Svensson A., *The ideal reporting interval for an epidemic to objectively interpret the epidemiological time, 2007 http:// rsif.royalsocietypublishing.org/content/7/43/297.full*

[12] P. Eckhoff, *Malaria parasite diversity and transmission intensity affect development of parasitological immunity in a mathematical model, 2012. http://www.malariajournal.com/content/11/1/419*

[13] L. J. S.Allen, *An Introduction to Stochastic Processes with Applications to Biology*, 2th Edition, Chapman and Hall, 2010.

NKEMNOLE, Edesiri Bridget received her masters degree in Statistics from University of Lagos, Nigeria in 2005 and Ph. D. in Statistics also in University of Lagos, Nigeria in 2012. She is currently a Lecturer I in University of Lagos, Nigeria. Her research interests are in the area of mathematical and Computational Statistics, Stochastic Processes and Time series analysis.
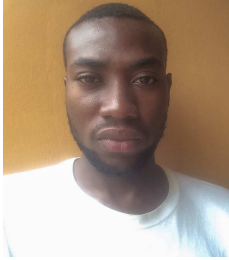
OSUNKEYE, Rapheal Opeoluwa received his Bachelor Degree in Mathematics/Statistics from University of Lagos, Nigeria in 2013. He is currently pursuing his masters degree in Statistics also in University of Lagos, Nigeria.