A mathematical model of sickle cell genome frequency in response to selective pressure from malaria

Chelsea Liddell, Nicci Owusu-Brackett, Dorothy Wallace *

September 22, 2011

Abstract

Because of the relative prevalence of hereditary sickle cell disease and the auxiliary role of the sickle cell gene in reducing the mortality of malaria, it is believed that *P. falciparum* has exerted selection pressure on human populations to increase the prevalence of this otherwise detrimental gene. A model incorporating three genotypes and two age cohorts is used to test the hypothesis that higher death rates due to malaria can exert selective pressure to increase the prevalence of the sickle cell gene. The model displays selection pressure for the carrier gene in the presence of increasing malaria death rates for either adults or children, showing both higher final frequencies of the gene as well as shortened time to reach these frequencies.

1 Introduction

Malaria is a mosquito borne disease endemic to tropical regions. Four species of malaria infect human beings, resulting in mild to acute illness depending on the species and the degree of

*Dartmouth College, Hanover, NH, USA

acquired resistance achieved by constant reinfection. One species, *Plasmodium falciparum*, is frequently fatal, particularly among children who have not yet built up significant immunity to the parasite. It is believed that *P. falciparum* is the most recently evolved human malaria species [4]. Malaria has been blamed not only for illness and death but also for lower economic growth in endemic regions, creating an economic as well as health burden [4], [6].

The hereditary disease of sickle cell is caused by a recessive gene that deforms red blood cells, resulting in severe anemia among homozygous individuals and mild symptoms among heterozygous carriers of the gene. Malaria, a parasite which during one part of its life cycle infects red blood cells, increases the tendency of infected cells to sickle, causing a very high mortality rate among homozygous individuals [10]. Unexpectedly, heterozygous individuals experience some protection from malaria infections. Aidoo *et al* [1], demonstrate reduced mortality and morbidity, Aluoch [3] reports higher resistance to malaria, whereas Hesran *et al* [7], demonstrate a reduced parasite load for heterozygous carriers of this otherwise damaging gene. Various explanations of the mechanism of this protection have been offered based on the observed tendency of red blood cells of heterozygous carriers to sickle under the stress of malaria, leading to their removal by the immune system [9] [11].

In 1954 Allison [2], noted that the distribution of the sickle cell trait seemed to correlate with malaria prevalence, with incidence of the sickle cell gene ranging from 15 to 45 percent in regions where malaria is present year round. Gallup & Sachs [6] report that almost all sickle cell homozygous children in developing countries die before reaching childbearing age. The relative prevalence of such a debilitating disease suggests that *P. falciparum* has exerted selection pressure on human populations to increase the prevalence of the sickle cell gene.

The model developed below is used to test the hypothesis that, given lower malaria death rates for heterozygotes, higher death rates due to malaria can exert selective pressure to increase the prevalence of the sickle cell gene. Results are compared to known prevalence rates.

2 Model Formulation

The model constructed after Irizarry *et al* [8], incorporates three sickle cell genotypes: heterozygous for the sickle cell gene (carrier), homozygous with no sickle cell gene (noncarrier) and homozygous for the sickle cell gene (afflicted). Because malaria results in higher morbidity and mortality in young children, the population is divided into children age 0-5 and adults (defined to be anybody over 5 years of age). It is assumed that only adults breed, creating an age segregated population for a total of six populations. The box model representing the system is shown in Figure 1.

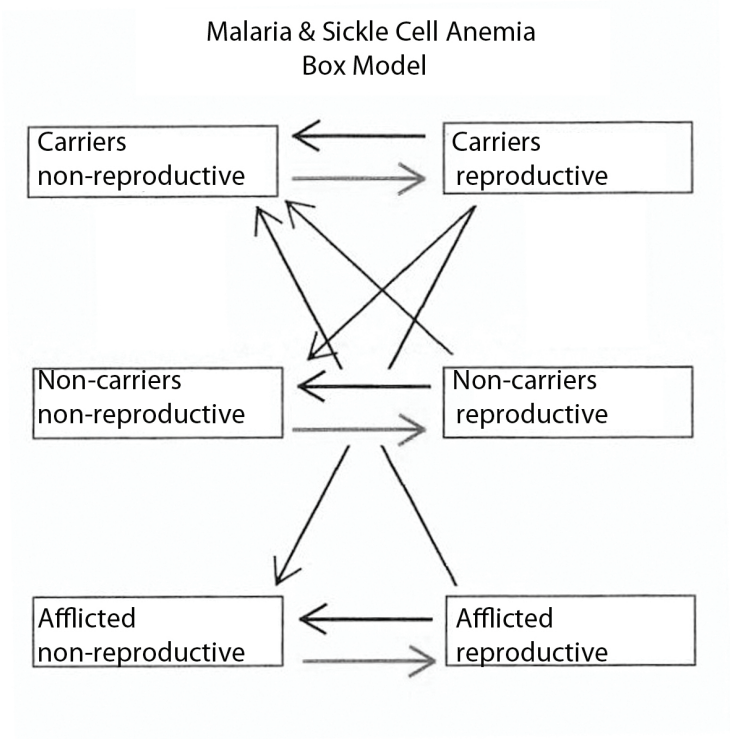

Figure 1: A schematic box model for the six populations in the model for sickle cell genotype growth. Boxes on the left represent the three genotypes in the population of children under 5. Boxes on the right represent the three genotypes of humans over 5 years of age. Black arrows represent reproduction and gray arrows represent maturation. Death rates are not represented in this picture.

Equations 2.1- 2.6 represent the model.

Change in carrier adults, Z = maturation - natural death.

$$Z' = aW - jZ \tag{2.1}$$

Change in noncarrier adults Y = maturation - natural death - malaria death.

$$Y' = aV - jY - kY \tag{2.2}$$

Change in afflicted adults X = maturation - natural death - sickle cell death.

$$X' = aU - jX - lX \tag{2.3}$$

Change in carrier children W = (growth) - maturation.

$$W' = m(\frac{(0.5)YZ}{(X+Y+Z)} + \frac{(0.5)Z^2}{(X+Y+Z)})(1-(U+V+W+X+Y+Z)) - aW \tag{2.4}$$

Change in noncarrier children V= (growth) - malaria death - maturation.

$$V' = m(\frac{Y^2}{(X+Y+Z)} + \frac{(0.5)YZ}{(X+Y+Z)} + \frac{(0.25)Z^2}{(X+Y+Z)})(1-(U+V+W+X+Y+Z)) - hV - aV \tag{2.5}$$

Change in afflicted children U =(growth) - maturation - sickle cell death.

$$U' = m(\frac{(0.25)Z^2}{(X+Y+Z)})(1-(U+V+W+X+Y+Z)) - aU - lU \tag{2.6}$$

Growth rates for children of each genotype were derived from a version of the logistic equation. The general population growth rate m is .04. This rate is multiplied by the total contribution from each genotype. For example noncarrier children may result from a union

of two noncarrier adults with probability jointly proportional to their respective frequencies:

$$\frac{Y^2}{(X+Y+Z)^2} \tag{2.7}$$

or from a union of one carrier and one noncarrier adult with probability

$$\frac{(0.5)YZ}{(X+Y+Z)^2} \tag{2.8}$$

or from a union of two noncarriers with probability

$$\frac{(0.25)Z^2}{(X+Y+Z)^2} \tag{2.9}$$

Together these quantities give a growth rate for the genotype which is multiplied by the total adult population, $(X+Y+Z)$ and a term that reflects a carrying capacity of one, $(1-(U+V+W+X+Y+Z))$. A carrying capacity is necessary in this problem to create the competition that hypothetically creates selection pressure for one genotype over another. The carrying capacity is assumed to include all "natural" death rates for children other than those specific to malaria or sickle cell. The capping term was included for the child populations. It could as easily have been put on all six populations or just on the adult population. Experiments with all three possibilities showed little difference in the overall pattern of response to malaria death rates. The variable represent percentages of the carrying capacity. Introducing an arbitrary carrying capacity and rescaling to percents results in the same equation. Thus, varying the carrying capacity does not affect any aspect of the dynamics as expressed in percent of that capacity.

3 Parameter Estimation

The *birth rate* is set to 4 percent per year, as estimated in [12]. The *rate of death due to sickle cell* is set at 50% per year, as estimated in [10]. Individuals afflicted with the homozygous sickle cell trait have high death rates and are assumed not to reproduce at all. Thus the quantity X never appears in any of the growth terms. The *maturation rate* from childhood to adult is 20% per year, based on five years of childhood. This rate is of course high, as only five year olds actually mature and other death rates would reduce the fraction of children at age 5. The cutoff from childhood to adult at this age is based on the large difference in malaria death rates between children and adults.

The *rate of death due to malaria* is varied for both children and adults. For different runs of this model, different values for child (under 5) and adult malaria death rates were used. The reported under-five malaria death rate is the most studied, and varies greatly based on region, and the year of reporting. The under-five malaria death rate ranges from 0 to 0.025 [13], [14]. However, historically the range may have been higher. Therefore, in this model, under-five death rates from 0 to 0.04 were included. Deaths from malaria for the population older than five years seem to be ignored in many studies, probably because this death rate is so small compared to the child death rates. When child death rates are higher, the reported adult death rate is even lower [14]. This model includes adult death rates from 0 to 0.0006, the highest rate found [14]. Because the source of variation in adult death rates in this model comes from death due to malaria and death due to sickle cell, the natural death rate was used as the only source of death for carriers (both children and adults), while adding an extra term for malaria to the noncarriers and an extra term for sickle cell death to the afflicted population. One should therefore interpret both malaria and sickle cell death rates to be the additional burden of these diseases on noncarrier and afflicted populations respectively, in comparison to their effect on carriers (which is included in the natural death

rate).

The *adult natural death rate* was set to .0004 (.04% per year) for all runs except those indicated. Actual reported death rates are much higher, up to one percent [5]. However the reported rates do not distinguish death due to malaria, which can range as high as four percent per year for the cohort under five years of age. For example, if the (prehistoric) life span is 30 and if the death rate due to malaria is .04 for ages 0-5, then that accounts for a death rate of .0067, (.67 percent per year, since that age group is one sixth of the population). This leaves .003 as an estimate of nonmalaria related death, most of which is in childhood. There is still some adult death due to malaria, as much as .0006 (.06 percent per year, [14]). This puts the total malaria death rate at .0073, leaving only .0027 residual death rate for a total of one percent. However, figures in this paper used the lower (.0004) rate, so that most runs could arrive at equilibrium in the 300,000 year duration of the simulation. Figure 2, which shows a typical run, is a particularly fast example, arriving at equilibrium in less than 100,000 years.

All quantities in the model are summarized in Table 1.

Table 1

quantity	variable	units	value	range	source
afflicted children	U	scaled			(computed)
noncarrier children	V	scaled			(computed)
carrier children	W	scaled			(computed)
afflicted adults	X	scaled			(computed)
noncarrier adults	Y	scaled			(computed)
carrier adults	Z	scaled			(computed)
birth	m	% /yr	.04		[12]
maturation	a	% /yr	.20		estimated
natural death	j	% /yr	.0004		[5]
death due to SC	l	% /yr	.50		[10]
child malaria death	h	% /yr	varies	0 to .04	[13],[14]
adult malaria death	k	% /yr	varies	0 to .0006	[14]

4 Analysis

The Jacobian at the sickle cell disease free equilibrium is computed in terms of Y^* and V^*, the values of Y and V at equilibrium. The actual values of Y^* and V^* are easy to compute, giving

$$Y^* = aj^{-1}V^* \tag{4.1}$$

and

$$V^* = maj(maj + ma^2 + hj^2 + aj^2) \tag{4.2}$$

for nonzero m, j, and a.

With entries in reverse alphabetical order, Z, Y, X, W, V, U, the Jacobian at the disease free equilibrium is given by:

$$\begin{pmatrix} -j & 0 & 0 & a & 0 & 0 \\ 0 & -j-k & 0 & 0 & a & 0 \\ 0 & 0 & -j-l & 0 & 0 & a \\ .5m & 0 & 0 & -a & 0 & 0 \\ .5mY^* + .5mV^* - .5m & m(1-2Y^*-V^*) & -m(1-V^*) & -mY^* & -mY^*-h-a & -mY^* \\ 0 & 0 & 0 & 0 & 0 & -a-l \end{pmatrix}.$$

Note that although the equations are nonlinear, the Jacobian at the disease free equilibrium simplifies after some algebra to the above expression.

A calculation shows the characteristic polynomial has the following factors:

$$(-a-l-\lambda) \tag{4.3}$$

$$(-j - l - \lambda) \tag{4.4}$$

$$((-j - k - \lambda)(-mY^* - h - a - \lambda) - a(m(1 - 2Y^* - V^*))) \tag{4.5}$$

$$((-j - \lambda)(-a - \lambda) - .5am) \tag{4.6}$$

The last factor gives a positive root when $j < .5m$. The parameter j was chosen to be quite small for the simulations in this paper, so the disease free equilibrium is unstable for all runs. This means that an arbitrarily small percentage of the sickle cell genotype into the population will not disappear, even though it increases the death rate among some subpopulations. Thus the sickle cell gene establishes a permanent presence in the population.

Note also that the populations all decline if their sum exceeds one. Thus any endemic equilibrium is in the $[0, 1]$ range. We do observe such an equilibrium in all simulations, as described below in the numerical results.

5 Numerical Results

In order to simulate this model in MATLAB, the system of differential equations ran to t=300,000, using time increments of one year. Initial values for all runs set child non-carriers at 0.4, adult non-carriers at 0.59, adult carriers at 0.01 and all other populations to 0. If the value for any population went below zero at any point, it was redefined to be exactly zero.

Runs were repeated for combinations for child malaria mortality (h) and adult malaria mortality (k) values, where h varied from 0 to 0.04 in steps of 0.0005 and k varied from 0 to 0.0006 in steps of 0.00005. In each run through the differential equation, the year at which

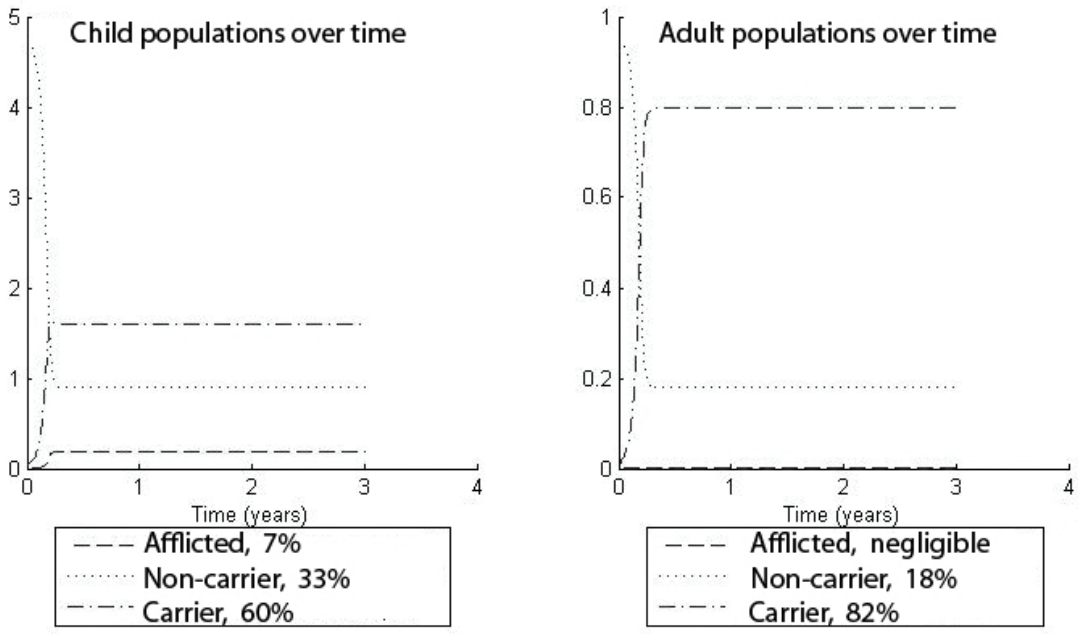

Figure 2: A typical run starting with one percent adult carriers. Due to the high death rate of sickle cell, there are almost no afflicted adults in the population at equilibrium. Child malaria mortality is set at .04 and adult malaria mortality is set at .0006. The horizontal scale is in 100,000 year intervals. Child populations are multiplied by .001. Approximate percents of total child and adult populations at equilibrium given in legend.

the populations reached 0.995 of their ultimate value was recorded as the time to steady state. The program identified a quantity to be at equilibrium when its difference quotient dropped below 0.005. The system was recorded as being at equilibrium when all quantities were at equilibrium.

Simulations with very low malaria death rates did not reach steady state by the end of the run. The cutoff points for this effect were identified by finding the parameters for which the populations barely reached steady state. These values of h and k are therefore not plotted in the figures. Finally, for each h and k, the proportion of each genotype in the population at equilibrium was calculated and recorded along with the time required to reach

equilibrium. An example of an individual run is shown in Figure 2. The overall results are summarized in Figures 3-5.

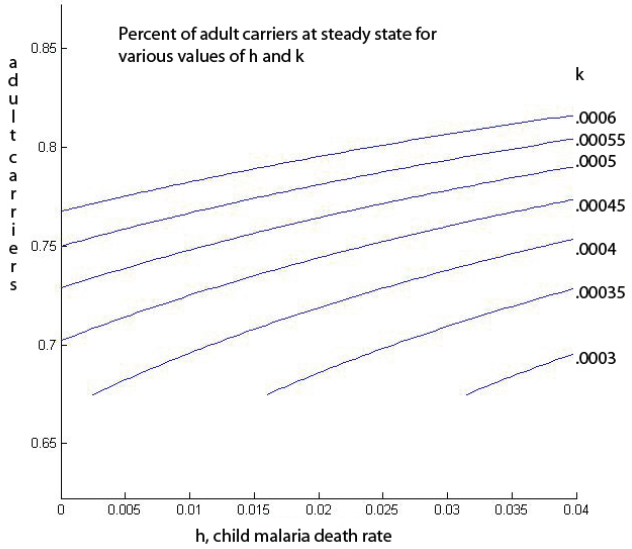

Figure 3: Fraction of adult carriers at equilibrium for various malaria death rates as h (child malaria death rate) and k (adult malaria death rate) vary.

Comparison of test runs across a range of malaria death rates confirms selection pressure in favor of the sickle cell gene. A typical run shown in Figure 2 shows the increase in carrier populations over time. As the malaria death rates of either child or adult population rise, the equilibrium frequency of the sickle cell gene also rises. Figure 3 shows the fraction of adult carriers at the endemic equilibrium increases with both child and adult malaria death rates. Figure 4 shows the fraction of child carriers at endemic equilibrium across the same range. Figure 5 shows the time required to reach steady state for various malaria death rates.

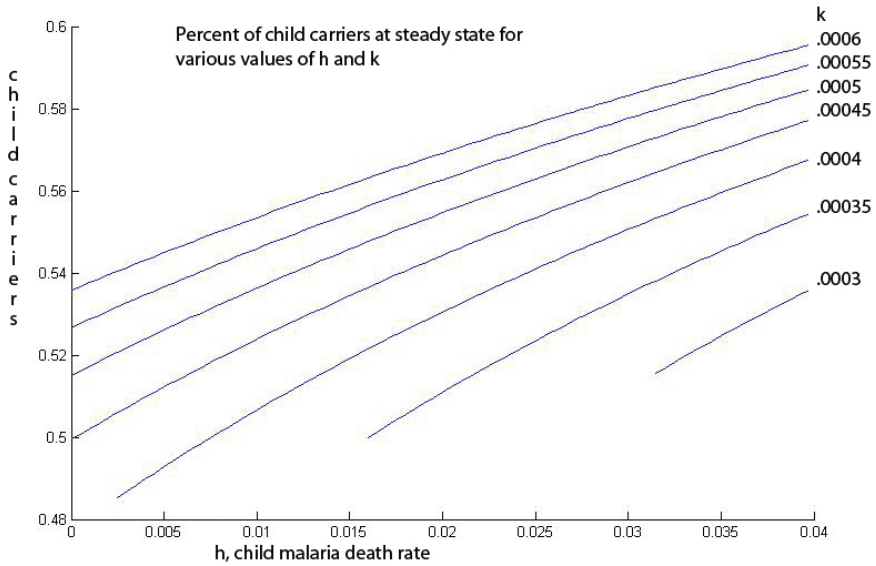

Figure 4: Fraction of child carriers at equilibrium for various malaria death rates as h and k vary.

Taken together, Figures 3-5 illustrate how the model displays selection pressure for the carrier gene in the presence of increasing malaria death rates for either adults or children, showing both higher final frequencies of the gene as well as shortened time to reach these frequencies.

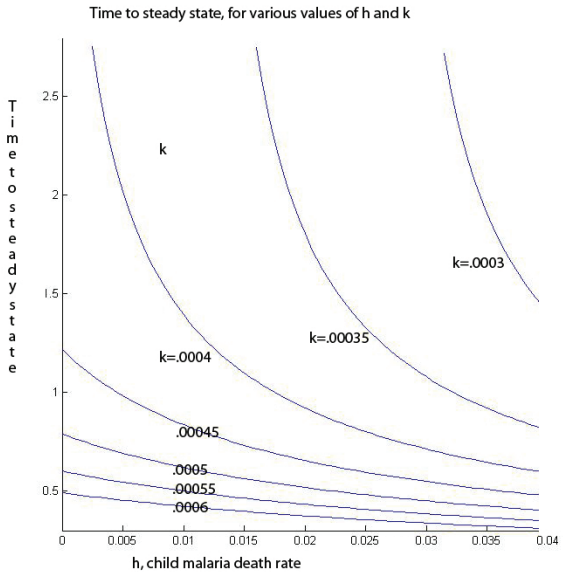

Figure 5: Time to steady state for various malaria death rates as h and k vary. Time units are in 100,000 years.

6 Discussion

The damaging sickle cell genotype is responsible for a far higher death rate among children and adults than malaria. Yet the protection against malaria enjoyed by heterozygous carriers of the gene has been proposed as source of selection favoring a high prevalence of this genotype in populations routinely exposed to malaria. The model studied in this paper expresses the tension between these two forces: the consistently high death rate and zero reproduction associated to homozygous carriers of sickle cell, the varying death rates for both children and adults due to malaria, and the relative protection against malaria for heterozygous carriers of the sickle cell gene. Our model supports the hypothesis that malaria protection selectively favors the sickle cell gene. In particular,

1. A small initial amount of the sickle cell genotype rises in the presence of malaria death

rates to an equilibrium far larger than the initial value would suggest.

2. Other factors being equal, increased adult malaria death rates results in larger equilibrium proportions of carrier populations and faster arrival at equilibrium.

3. Other factors being equal, increased child malaria death rates results in larger equilibrium proportions of carrier populations and faster arrival at equilibrium.

Additionally the model supports the idea that *P. falciparum* is a recently evolved organism relative to 200,000 to 500,000 years of human history. The equilibrium values for carrier populations reached by our models were above 60 %, which is consistently higher than those observed in the literature, suggesting that equilibrium has not been achieved yet.

References

[1] Aidoo, M., Terlouw, D.J., Kolczak, M.S., McElroy, P. D., ter Kuile, F.O., Kariuki, S., Nahlen, B.L., Lal, A.A., Udhayakumar, V. (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality, *Lancet.*, *359*, 1311-12.

[2] Allison, A.C. (1954). The distribution of the sickle-cell trait in East Africa and elsewhere and its apparent relationship to the incidence of subtertian malaria, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *48*(4), 312-318.

[3] Aluoch, J.R.(1997). Higher resistance to *Plasmodium falciparum* infection in patients with homozygous sickle cell disease in western Kenya, *Tropical Medicine and International Health*, *2*(6), 568-571.

[4] Bremen, J.G. (2001). The ears of the hippopotamus: manifestations, determinants and estimates of the malaria burden, *Am. J. Trop. Med. Hyg.*, *64*, 1-11.

[5] CIA online factbook, retrieved from *https://www.cia.gov/library/publications/the-world-factbook/fields/2066.html*

[6] Gallup, J.L., and Sachs, J.D. (2001). The Economic Burden of Malaria, *Am. J. Trop. Med. Hyg.*, *64*, 85-96.

[7] Hesran, J.Y., Personne, I., Personne, P., Fievet, N., Dubois, B., Beyemé, M. Boudin, C., Cot, M., and Deloron, P. (1999). Longitudinal study of *Plasmodium falciparum* infection and immune responses in infants with or without the sickle cell trait, *International Journal of Epidemiology*, *28*, 793-798.

[8] Irizarry, K., Fish, M., Madigan, D. and Taub, J. (2008). Shelter from the Swarm: The Malaria-Sickle Cell Connection *Dartmouth Undergraduate Journal of Science*, *4*, 28-32.

[9] Luzzato, L. (1970). Increased sickling of parasitised erythrocytes as mechanism of resistance against malaria in the sickle cell trait, *Lancet*, *295* (7642), 319-22.

[10] Makani,J., Komba, A.N., Cox, S.E., Oruo, J., Mwamtemi, K., Kitundu, J., Magesa,P., Rwezaula, S., Meda, E., Mgaya, J., Pallangyo, K., Okiro, E., Muturi, D., Newton, C.R., Fegan, G., Marsh, K., and Williams, T.N. (2010). Malaria in patients with sickle cell anemia: burden, risk factors, and outcome at the outpatient clinic and during hospitalization, *Blood*, *115*(2), , 215-220.

[11] Nagel, R., and Roth, E.F. (1989). Malaria and red cell genetic defects,*Blood*, *74*, 1213-21.

[12] Population Reference Bureau, retrieved from *http://www.prb.org/Datafinder/Topic/Bar.aspx?sort=vo*

[13] Rowe, A.K., Rowe, S.Y., Snow, R.W., Korenromp,E.L., Armstrong Schellenberg, J.R., Stein, C.,...Steketee, R.W. (2006). The burden of malaria mortality among African children in the year 2000. *International Journal of Epidemiology*, *35*, 691-704.

[14] World Bank. (2006). Disease and mortality in Sub-Saharan Africa. World Bank Publications, 195-213.