



472

2015

Malaria and Early African Development: Evidence from the Sickle Cell Trait

Emilio Depetris-Chauvin, David N. Weil

Malaria and Early African Development: Evidence from the Sickle Cell Trait*

Emilio Depetris-Chauvin[†]

David N. Weil[‡]

Abstract

We examine the effect of malaria on economic development in Africa over the very long run. Using data on the prevalence of the mutation that causes sickle cell disease, we measure the impact of malaria on mortality in Africa prior to the period in which formal data were collected. Our estimate is that in the more afflicted regions, malaria lowered the probability of surviving to adulthood by about ten percentage points, which is roughly twice the current burden of the disease. The reduction in malaria mortality has been roughly equal to the reduction in other causes of mortality. We then ask whether the estimated burden of malaria had an effect on economic development in the period before European contact. Using data at the ethnic group level, we find little evidence of a negative relationship between malaria burden and population density or other measures of development. We also construct a simple macro-demographic model to assess the expected effect of observed malaria mortality, and find that it to be small.

1 Introduction

It is impossible to understand the pattern of comparative economic development in the world today without understanding comparative development in the past. Consider, for example, a horizon of 500 years. Countries and regions that were highly developed as of the year 1500 are, for the most part, among the most developed today. Exceptions to this regularity, such as China, tend to be growing quickly. Taking into account flows of population over the last

*We are grateful to Ronald D. Lee, Andrew Mason, Gordon McCord, Nathan Nunn, Marcella Alsan, and Fred Piel for sharing data, to Fatima Aqeel, Federico Droller, Daniel Prinz, and Scott Weiner for research assistance, and to Quamrul Ashraf, Adriana Lleras-Muney, Stelios Michalopoulos and seminar participants at Boston University, Harvard University, Harvard Center for Population and Development Studies, Indiana University, New York University, Pennsylvania State University, Princeton University, UCLA, University of Connecticut, and University of Pennsylvania for helpful comments

[†]Universidad de Los Andes, Department of Economics. Email: e.depetris@uniandes.edu.co

[‡]Brown University and NBER. Email: davidweil@brown.edu

half millennium makes this correlation even stronger: Countries populated by people whose ancestors lived in the most developed countries are most likely to be rich today. Looking within countries, people who are descended from parts of the world that were highly developed in the year 1500 are on average higher up in the income distribution than people descended from regions that were not developed.¹ Going back further back in time, there is still strong predictive power of past development for present outcomes. Comin, Easterly, and Gong (2010) show that not only is a country's level of technology from 500 years ago predictive of income today, but so is the level of technology 2,000 or 3,000 years ago. Hibbs and Olsson (2004) show that the date at which the transition from hunting and gathering to settled agriculture took place is predictive of a country's income today.

The fact that development in the past is so predictive of development today suggests two possible theories. First, it may be that the same factors that influenced development in previous historical eras are still operative in the present. Examples of such factors are genetic attributes of populations, slowly changing aspects of culture or institutions, and characteristics of geography or climate.² Alternatively, it may be that the specific factors that caused past underdevelopment are no longer relevant today, but that the fact of past underdevelopment itself is causal of current underdevelopment. For example, it could be that the early development advantage of the Eurasian land mass arose from the historical presence of plentiful species of large seeded grasses and domesticable animals, as argued by Diamond (1997), but that the continuation of the development gap between Eurasia and other regions results from the effect of colonial institutions that Europeans were able to impose on much of the rest of the world as a result of this initial advantage.³ A related argument, stressed by Spolaore and Wacziarg (2013) is that past differences in development are causal for current outcomes because of barriers to transmission of productivity-enhancing innovations among populations with different historical roots.

Whichever of these theories is correct (and obviously it is possible for both of them to have some validity), there is clearly much to be learned by looking at the roots of development differences in the past. In this paper, we examine the historical impact on development of malaria, which is one of the most significant diseases in the world today in terms of its humanitarian burden. Malaria's control is widely studied by biologists and social scientists. Economists actively debate its role in affecting growth in the modern world.⁴ However, as the above discussion makes clear, it would be possible that even if malaria were not important in affecting economic development today directly, it could nonetheless have been an important determinant of development historically and via that

¹Chanda and Putterman (2007), Putterman and Weil (2010).

²See, for example, Ashraf and Galor (2012) and Sachs, Malaney, and Spielman (2004). Spolaore and Wacziarg (2013) discuss the different possible channels by which characteristics that impact economic outcomes might be transmitted intergenerationally.

³See, for example, Acemoglu, Johnson, and Robinson (2001) and Nunn (2008).

⁴The widely quoted estimate of Gallup and Sachs (2001) is that malaria reduces growth of GDP per capita by 1.3 percent per year in the African countries most afflicted. See Weil (2010) for an extensive critique of this literature.

channel indirectly affect development today.

In studying the role of malaria in long run development, we are also inevitably studying the long run development of Africa. Both historically and today, Africa has been the major focus of the disease. Indeed, malaria was not present in the tropical regions of the new world until it was accidentally brought there by Europeans (McNeill, 1977). Historians of Africa attribute a large role to diseases in general, and malaria in particular, in shaping development (Akyeampong, 2006). For example, Webb (2006) describes malaria, along with trypanosomiasis (transmitted by the tsetse fly) as having profoundly influenced African patterns of settlement as well as culture (Alsan, 2014, also finds a large role for trypanosomiasis in shaping population density in Africa). The data we analyze below gives us a unique opportunity to assess the historical role of malaria in shaping geographic heterogeneity in Africa. Chiovelli, Cervellati, and Esposito (2015) claim that long term malaria exposure affected modern ethnic diversity, in particular contributing to more frequent endogamic marriages and stronger ethnic identities.⁵

Analysis of the role played by malaria in shaping historical development is severely hampered by a lack of data. Biologists only came to understand the nature of the disease in the late nineteenth century. Accounts from travelers and other historical records provide some evidence of the impact of malaria going back millennia, but these are hardly sufficient to draw firm conclusions.⁶ Even today, trained medical personnel have trouble distinguishing between malaria and other diseases without the use of microscopy or diagnostic tests. As discussed below, there do exist data (malaria ecology) that measure the extent to which the environment in different geographical regions is supportive of malaria. One can look at the empirical relationship between this malaria ecology measure and economic development, either currently or historically. However, such an approach faces severe limitations. One problem is that the malaria ecology variable is not scaled in way that allows for easy economic interpretation. Further, it is difficult to know whether one has controlled for correlates of malaria ecology that might independently influence the process of development. These correlates could be other factors that directly influence output today (either other diseases, or the tropical climate, which affects agriculture) or they could be the result of historical processes, for example institutional quality, which Acemoglu, Johnson, and Robinson (2001) argue was influenced by the disease

⁵Weil (2014) paints a picture of African development in 1500, both relative to the rest of the world and heterogeneity within the continent itself, using as his indicators population density, urbanization, technological advancement, and political development. Ignoring North Africa, which was generally part of the Mediterranean world, the highest levels of development by many indicators are found in Ethiopia and in the broad swathe of West African countries running from Cameroon and Nigeria eastward along the coast and the Niger river. In this latter region, the available measures show a level of development just below or sometimes equal to that in the belt of Eurasia running from Japan and China, through South Asia and the Middle East, into Europe. Depending on the index used, West Africa was above or below the level of development in the Northern Andes and Mexico. Much of the rest of Africa was at a significantly lower level of development, although still more advanced than the bulk of the Americas or Australia.

⁶Akyeampong (2006), Mabogunje and Richards (1985)

environment.

In this paper we address the lack of information on malaria’s impact historically by using genetic data. In the worst afflicted areas, malaria left an imprint on the human genome that can be read today. Specifically, we look at the prevalence of the gene that causes sickle cell disease. Carrying one copy of this gene provided individuals with a significant level of protection against malaria, but people who carried two copies of the gene died before reaching reproductive age (Hedrick, 2011). Thus the degree of selective pressure exerted by malaria determined the equilibrium prevalence of the gene in the population. By measuring the prevalence of the gene in modern populations, we can thus back out estimates of the severity of malaria historically. We compare these estimates to the burden of malaria today and discuss how the change in malaria mortality has compared to the change in mortality from other sources.

With estimates of the extent of malaria mortality in hand, we then turn to look at the impact of the disease on economic development. We take two approaches. The first is statistical: we present regressions of a number of measures of development within Africa on a malaria burden measure that we create based on sickle cell prevalence. Of particular note, we apply our analysis to a data set measured at the level of ethnic groups as an alternative to more common country-level analyses. We present simple OLS results, as well as results in which we instrument for malaria burden using an index of climactic suitability for malaria transmission. The result of this statistical exercise is that we find no evidence of malaria burden negatively affecting historical economic development. Our second approach eschews regressions, instead using the machinery of standard economic modeling, along the lines of Ashraf, Lester, and Weil (2007). In the current paper, we focus in particular on quantifying the economic effect of malaria mortality of the magnitude estimated from our genetic data. Our model predicts that the effect would be small.

The rest of this paper is organized as follows. In Section 2, we discuss the biology of the link between malaria and sickle cell disease. Section 3 presents and applies our model for using the current level of sickle cell prevalence to estimate the historical burden of malaria, including comparisons of malaria’s historical burden to its current level. We then turn to the question of how malaria affected development. Section 4 presents a statistical analysis of the relationship between the malaria burden measure we create and a number of measures of development within Africa. In Section 5 we take a model-based approach to evaluating cost of the malaria mortality, and also briefly discuss the effect of malarial morbidity. Section 6 concludes.

2 Malaria and Sickle Cell Disease

2.1 Background

Malaria is caused by the plasmodium parasite, which is transmitted to humans through the bite of a female anopheles mosquito. Early symptoms of malaria

include fever, chills, severe headache, and vomiting. In severe cases these are followed by respiratory distress, severe anemia, or neurological symptoms. Infants are protected from the disease in the first few months of life by a combination of maternal antibodies and characteristics of the structure of fetal hemoglobin. In malaria endemic areas, most children have developed substantial immunity by the age of five.

Africa currently accounts for 82 percent of world malaria cases and 90 percent of world malaria deaths (World Health Organization, 2014). The geographical pattern of malaria's severity is largely determined by the climactic conditions that support mosquito breeding as well as by the mix of mosquito species present. There are significant differences in the vectorial capacity among the approximately 20 species of anopheles that transmit malaria, based on factors such as the mosquito's preferred targets, biting frequency, and lifespan. The most effective vector, *Anopheles Gambiae*, is the principal vector in Africa.

Several mutations have arisen in human populations that provide resistance to malaria. These include the mutation causing thalassemia, which is present in Mediterranean, Arab, and Asian populations; the absence of the Duffy blood group in west Africa; hemoglobin E in Southeast Asia; and hemoglobin C in West Africa. The most important such mutation is the one that causes sickle cell disease (Allison, 2002; Kwiatkowski, 2005; Nelson and Williams, 2006).

The sickle cell trait is a mutation in the gene that produces hemoglobin, the oxygen-carrying component in red blood cells. Individuals carry two copies of this gene, one received from each parent. Individuals who carry one normal copy of the gene (referred to as A type) and one copy with the sickle cell mutation (S type) are carriers of the disease. In individuals of the AS genotype, a fraction of the hemoglobin in their red blood cells have an abnormal structure. In individuals who have two copies of the sickle cell gene (SS genotype), almost all hemoglobin molecules are of the abnormal type.

In conditions of inadequate oxygen supply (hypoxia), hemoglobin produced by the S gene becomes rigid, leading to a characteristic sickle shape of red blood cells. Carriers of sickle cell trait generally do not suffer many adverse effects.⁷ However, there can be negative consequences from sickling in conditions of low oxygen such as unpressurized airplane flights and extremely rigorous exercise (Motulsky, 1964). In individuals of the SS genotype, such sickling of red blood cells is far more common, leading to acute episodes of disease in which abnormally shaped cells restrict blood flow to organs. Such individuals also suffer from anemia and reduced resistance to infection. In 1994, life expectancy for SS children in the United States was 42 years for males and 48 years for females. In the absence of modern medical care, individuals of the SS genotype are not able to survive to adulthood.

The sickle cell mutation is relevant to malaria because infection of a red blood cell with the malaria parasite leads to hypoxia. In individuals of the AS genotype such blood cells sickle and are then eliminated by macrophage

⁷Williams et al. (2005) show the absence of any significant effect of carrier status on a wide range of childhood diseases.

cells of the body’s immune system, lessening the burden of infection (Luzatto, 2012). Carriers of the sickle cell trait are particularly resistant to severe malarial episodes; they are less resistant to mild cases. The mechanism by which carriers are protected from malaria is different than the acquired immunity that both AA and AS individuals achieve following repeated exposure to the disease.

The benefit that possessing a single copy of the sickle cell gene conveys counterbalances the biological cost incurred when homozygous SS children are stricken with sickle cell disease. An individual of the AS genotype is more likely to reach adulthood than is an individual of the AA genotype, but the former is also more likely to see his/her child die of sickle cell disease. This is known as a heterozygote advantage or balanced polymorphism. As shown more formally below, the stronger is the pressure of malaria on survival, the more advantaged are individuals who carry the S gene, and in equilibrium, the higher the percentage of the population who will be carriers. Indeed, it was the correlation of high prevalence of the sickle cell gene and the presence of malaria that first led scientists to understand the protective role of the sickle cell mutation (Kwaitkowski, 2005). As will be seen in the next section, the underlying genetic mechanism by which the sickle cell trait is transmitted provides a means of mapping sickle cell prevalence into an estimate of the mortality burden of malaria.

Piel et al. (2010) present a global geo-database of S gene frequency based on comprehensive electronic search of academic publications presenting S gene frequency figures. Each reference included in the dataset meets the criterion that the surveyed population was representative of the indigenous population of a particular location. Piel et al. assign a geographic coordinate to all samples with the distribution of AS and AA genotypes that meet their strict inclusion criteria. Using a Bayesian model-based geostatistical framework they then create a continuous map of the sickle cell gene frequency resulting in 10 km by 10 km resolution global raster grid. It is important to note that throughout our paper we use the terminology sickle cell trait prevalence (the fraction of people who are carriers of the allele S) instead of sickle cell gene (or S allele) frequency as in Piel et al (2010). Since very few S homozygotes survive to adulthood, the sickle cell trait prevalence is very close to twice of the sickle cell gene frequency.

Figure 1 presents the geographic distribution of S gene frequency from Piel et al. The maximum levels of the mutation are located in West-Central Africa (Gabon and Congo). The west part of Africa, from Cameroon to Senegal, shows a large heterogeneity in the prevalence of the mutation, ranging from almost complete absence in some places and very high values (above 15 percent) in others. Medium level frequencies are also located in the proximity of Lake Victoria. Finally, no mutation is documented in the southern part of the continent.⁸

⁸Genetic evidence supports the hypothesis that the sickle cell trait observed today results from five independent mutations, four in Africa and one in Asia. The four African haplotypes are Benin, Senegal, Cameroon, and Bantu (Central Africa). Uncertainty exists regarding the dates of these mutations. Some scholars have argued that they probably arose fewer than 6,000 years ago in response to the spread of agriculture (see, for example Weissenfeld, 1967 and Livingstone, 1958). Molecular genetic analyses have yielded a broad range of

3 Measuring the Historical Burden of Malaria Using Data on the Sickle Cell Trait

3.1 Model

Our goal is to examine what the prevalence of the sickle cell trait among African populations tells us about the impact of malaria historically. As described above, every adult carries two alleles. These can be either sickle cell (S) or normal (A). A person with the SS genotype will develop sickle cell disease and not survive to adulthood. A person who carries the sickle cell trait (AS) will have a survival advantage against malaria in comparison to someone who doesn't (AA).⁹

We consider a simple model in which deaths occur due to either malaria or to other causes. Let M be the probability of dying of malaria, conditional on not dying of something else. Similarly, let P be the probability of dying of something else, conditional on not dying from malaria.¹⁰ The deaths that we are concerned with are those between birth and adulthood, which is taken to be the age at which children are produced. The number of adults from a cohort of newborns will be given by:

$$\text{Surviving Adults} = (1 - M)(1 - P)\text{Newborns} \quad (1)$$

Throughout the analysis, we will assume that the probability of dying from non-malaria causes, P , is the same for individuals with the AS and AA genotypes. The probabilities of dying from malaria, differ between AA and AS genotypes. We designate these probabilities M^{AA} and M^{AS} . It is also useful to designate the relative survival rates of these two genotypes, which we call β :

$$\beta = \frac{(1 - M^{AA})(1 - P)}{(1 - M^{AS})(1 - P)} = \frac{(1 - M^{AA})}{(1 - M^{AS})} \quad (2)$$

β is the probability of a non-carrier living into adulthood relative to the probability of a carrier living into adulthood. The smaller is β , the larger is the advantage of the AS genotype. A value of $\beta = 1$ would indicate that there is no advantage to carrying the sickle cell gene. The values for M^{AA} and M^{AS} , and thus for β , will depend on both the disease environment and the state of medical technology. For example, in a place where there are no malaria mosquitoes, M^{AA} and M^{AS} will both be equal to zero, and β will be equal to one. Clearly, the availability of modern medical care should mean that both

estimates, ranging from between 3,000 and 6,000 generations ago (Kurnit, 1979 and Solomon, 1979) to, in the case of the Senegal haplotype, less than 3,000 years (Curat et al, 2002).

⁹We ignore other mutations. In the presence of other mutations that also control malaria, the genetic benefit of carrying the S trait is reduced, but its cost remains the same. Thus our analysis, based only on the S trait, will understate the full burden of malaria.

¹⁰For the purposes of this part of the analysis, it does not matter whether deaths due to sickle cell disease are counted as being due to malaria or due to a non-malaria cause. In our analysis below, we include deaths due to sickle cell disease as part of our measure of malaria burden, because in the absence of the disease, such deaths would not occur.

M^{AA} and M^{AS} are lower today than they were in the past. However, it is not clear a priori which mortality rate would be reduced by more.

Relative Survival in Modern Populations

Although our main objective is asking what role malaria played historically, it is of interest to see what information is available about relative survival today. One way to measure relative survival is to compare prevalence among adults to that among newborns. The relevant equation is:

$$\frac{AA\ Adults}{AS\ Adults} = \frac{(1 - M^{AA})(1 - P)AA\ newborns}{(1 - M^{AS})(1 - P)AS\ newborns} \quad (3)$$

Rearranging the equation:

$$\beta = \frac{(1 - M^{AA})}{(1 - M^{AS})} = \frac{AA\ Adults}{AS\ Adults} \times \frac{AS\ newborns}{AA\ newborns} \quad (4)$$

Morrow and Moss (2006) report on data from the Garki study, conducted in a region of high malaria transmission in Nigeria in the 1970s. Among adults, 29.0 percent of the population was AS, while 70.2 percent were AA. Among newborns, 23.6 percent were AS and 73.8 percent AA. Entering these figures in the equation above gives a value of $\beta = .775$ percent. More generally, Morrow and Moss report that the prevalence of the sickle cell trait in West Africa trait rises from 20-24 percent among newborns to 26-29 percent among adults.¹¹

Another approach is to look at survival directly. Motulsky (1964, table 2) examined the relative survival of over 15,000 children in the Congo in the 1950s, a time and place where modern treatments for malaria would have been relatively scarce. The study compared families where one parent was AS and one AA, on the one hand, to families in which both parents were AA, on the other. Since half of the children in the former group were carriers, compared to none in the latter, one can back out the relative survival of AS vs. AA children. The study found mortality from all causes of 24.0 percent among AS children vs. 27.4 percent among AA. The implied value of β is 0.955. Part of the explanation for the different estimates of β in the Congo vs. West Africa may be a difference in the severity of malaria. In the Congo, the malaria ecology index (discussed below) is 12.1; in West Africa it is generally in the neighborhood of 20.

3.2 Measuring Relative Survival in Historical Populations

We now turn to our main line of inquiry, which is using observed frequency of the sickle cell trait to back out the severity of malaria. Let π_t be the fraction of

¹¹Examining the Baamba of Uganda, who live in extremely high malaria environment, Lehmann and Raper (1956) report that the fraction of sicklers (AS or SS) rises from 30 percent of those under five years age to 37 percent among adults. Using these figures in equation (4), this would imply a value of $\beta = 73$ percent. This calculation understates the size of the mortality differential, however, both because those in the under five age group will have already experienced differential mortality from malaria and because the group of young sicklers presumably includes a larger fraction of SS individuals than does the adult group.

adults in generation t who are carriers (AS). We assume that no one born with SS lives into adulthood. Thus the fraction of the adult population who are not carriers is $(1 - \pi_t)$. The fraction of alleles in the adult generation that are S is simply $\frac{\pi_t}{2}$. Assuming that mating between carriers and non-carriers is random, the fractions of children born who are of each type are:

$$\begin{aligned} AA &: \left(1 - \frac{\pi_t}{2}\right)^2 \\ AS &: \pi_t \left(1 - \frac{\pi_t}{2}\right) \\ SS &: \left(\frac{\pi_t}{2}\right)^2 \end{aligned}$$

The difference equation for π , which relates prevalence among adults in successive generations, is:

$$\pi_{t+1} = \frac{\pi_t \left(1 - \frac{\pi_t}{2}\right)}{\pi_t \left(1 - \frac{\pi_t}{2}\right) + \beta \left(1 - \frac{\pi_t}{2}\right)^2} = \frac{\pi_t}{\beta + \pi_t \left(1 - \frac{\beta}{2}\right)} \quad (5)$$

We solve for the steady state by setting $\pi_t = \pi_{t+1}$:

$$\pi_{ss} = \frac{1 - \beta}{1 - \frac{\beta}{2}} \quad (6)$$

This has the properties we would expect: the smaller is the β , that is the greater is the survival advantage of being a carrier, the larger is the equilibrium fraction of the adult population that will be carriers.

We can turn this equation around to infer the burden of malaria on survival based on the prevalence of the sickle cell trait among adults:

$$\beta = \frac{1 - \pi_{ss}}{1 - \frac{\pi_{ss}}{2}} \quad (7)$$

This equation says that if 20 percent of the adult population are carriers in a steady state, then $\beta = 0.89$, in other words non-carriers are only 89 percent as likely to live to adulthood as carriers.

Current vs. Historical Prevalence.

The analysis above considers a population that is in equilibrium in terms of the selective impact of malaria and the prevalence of the sickle cell trait. Such a steady state presumably existed in Africa in the period before European contact. However, the only data on sickle cell prevalence available comes from observations over the last 70 years.¹² One might worry that modern prevalence rates are not the same as those that held historically, because the health environment has been changing over time. To address this question, it is straightforward to

¹²Testing predates modern technology for genetic analysis. Carriers of the sickle cell gene can be reliably diagnosed by taking a drop of blood and mixing in a reducing agent to induce hypoxia, then examining with a microscope whether blood cells have sickled.

use equation (5) to look at how prevalence π changes in response to a change in relative survival β .

As an example, we consider the case where there is initially a steady state of $\pi = .20$, and correspondingly $\beta = 0.89$. In generation 1, the value of β is set to one, corresponding to the complete eradication of malaria, or the removal of the population to a place where the disease is not present. Figure 2 shows the fraction of the population that will be carriers of the sickle cell trait after a set number of generations. The figure shows that the initial decline is very rapid, but that there is long tailing off once the prevalence gets sufficiently low.

In principle, one could test this model by looking at how the prevalence of the sickle cell trait changes in a population that was removed from exposure to malaria at a known time. Although the data are in practice not sufficient for a formal test, the case of African Americans provides a rough consistency check. We start by calculating the fraction of adults who are carriers in the countries that were the source of slaves that came to the United States. Specifically, we combined two data sources. We used data from Piel et al. (2010) on the prevalence of the sickle cell trait among adults in modern African countries. We combined these with estimates from Putterman and Weil (2010, main appendix, part II.3) on the fractions of the slaves who came to the United States originating in each African country. Multiplying and summing these two series, we estimate that 16.2 percent of slaves who came to the United States were carriers of the sickle cell trait (calculation available on request).

The other piece of information needed is the date on which the ancestors of today's African Americans left Africa. Obviously, this took place over several centuries, and we do not know of an estimate of the average date of departure. As a rough estimate, we assume that the average ancestor of today's African Americans left Africa around 1750 – a span of roughly 10 generations.

Starting with 16.2 percent prevalence and applying equation (5) for ten generations would lead to a prevalence of 9.0 percent today. However, one has to deal with the issue of admixture with other populations. Putterman and Weil (2010, Appendix B) summarize literature on the fraction of African American heritage that is due to non-Africans, reporting 20 percent as a consensus figure. That admixture took place at unknown points in time. Mechanically, the later that admixture took place, the lower would be the prevalence of the sickle cell trait among African Americans today. If the entire admixture took place in the last generation, the prevalence today would be 7.2 percent. By contrast, if the entire admixture took place ten generations ago, our calculation would yield a prevalence of 7.9 percent today.

In fact, the prevalence of the trait among African Americans today is 8 percent (National Heart, Lung, and Blood Institute 1996). Thus the model slightly under-predicts the prevalence of the sickle cell trait among African Americans. One possible source of this error could be that that African slaves brought to the Americas did not find themselves in a completely malaria-free environment. For example, McGuire and Coelho (2011) stress the immunity of slaves to malaria as one of the reasons that southern planters favored them over indentured servants.

In terms of our use of current prevalence of the sickle cell trait to measure

the historical burden of malaria, it is not clear that this analysis of the dynamics of prevalence matters much, since there is little reason to believe that contact with Europeans did anything to reduce the impact of malaria in Africa until the second half of the twentieth century, which is when most of the measures of sickle cell come from.

3.3 Measuring the Overall Burden of Malaria

Knowing the relative survival of carriers and non-carriers does not tell us the overall effect of malaria, for two reasons. First, in addition to deaths from malaria, we must take into account the cost of sickle-cell disease itself. This issue is easily addressed by looking at the fraction of adults who are carriers, from which we can derive the fraction of children who will suffer from sickle cell disease. The underlying data on the number of carriers is the same data used to estimate β .

The second reason that knowing β (relative survival) does not tell us the overall burden of malaria is because a given value of β could be consistent with different levels of absolute survival. For example, $\beta = 0.80$ could be consistent with $M^{AA} = 0.20$ and $M^{AS} = 0$, but it could also be consistent with $M^{AA} = 0.60$ and $M^{AS} = 0.50$. Dealing with this issue requires bringing to bear additional data. Specifically, we need an additional piece of information on M^{AA} , M^{AS} , or their ratio. One can look at modern populations for some information, with the caveat that modern data on survival is not necessarily informative about survival in Africa prior to European contact, where both the disease environment and the level of medical care differed from today.

Allison (2002, Table 2) reports results from an examination of 104 child malaria deaths from different countries in Africa, in which the weighted average prevalence of the sickle cell trait was 21 percent. Only one child examined had the sickle cell trait, which would suggest that the trait is almost completely protective against malaria death. However, a different set of investigations (Allison 2002, Table 1) that looked at severe *P. falciparum* infections rather than deaths found a relative incidence of infections in AS that was 46 percent as high that for AA. Both sets of studies just described were conducted in the 1950s or very early 1960s. A larger and more recent study (Hill et al., 1991) examined children in The Gambia. Children who were severely ill with malaria were compared to a control group. The severely ill children had cerebral malaria or severe malarial anemia. Without treatment, most of the children in this group would have died. Among the severe malaria group, the frequency of the AS genotype was 1.2 percent, while among the control group it was 12.9 percent. This implies that the relative risk of developing severe symptoms (and presumably dying without medical care) in AS as compared to AA is 0.08.¹³ A third study (Williams et

¹³Another study (Greenwood, Marsh, and Snow, 1991) examined children in Kenya, finding that the sickle cell trait was present in only 1.8 percent of children with severe malaria anemia but 3.9 percent of children with uncomplicated malaria. This finding confirms that the trait is more protective against severe malaria than against mild cases. However, because no data are given on the prevalence of the trait in the overall population, one cannot back out the

al, 2005) concluded that AS was 90 percent protective against severe or complicated malaria. These studies suggest that reasonable bounds on M^{AS} are zero on the low end, and to assume that $\frac{M^{AS}}{M^{AA}} = 0.08$ on the upper end.¹⁴

With estimates of M^{AS} and M^{AA} , we are in a position to look at the overall costs of malaria. There are three components to this cost: deaths from malaria among children who are carriers of the sickle cell trait, death from malaria among children who are not carriers, and deaths from sickle cell disease. Table 1 shows the fractions of births that fall into each category, the death rate for each group, and the total fraction of child deaths (from malaria or sickle cell disease) that are due to each category. The overall fraction of children who die due to malaria and sickle cell disease is simply the sum of the three terms in the right hand column.

Figure 3 does a more extensive analysis, considering different values of π , the prevalence of the sickle cell trait among the adult population. We consider values ranging from zero to 40 percent, which is the highest level observed among specific populations. For each value of π , we calculate the implied value of β , assuming that the prevalence represents a steady state (equation 7). We also show the fraction of newborn children who will die of sickle cell disease and the malaria death rates for non-carriers under the two different assumptions about the death rate for carriers discussed above, specifically that $M^{AS} = 0$ and that $\frac{M^{AS}}{M^{AA}} = 0.08$. Finally, we show the overall burden of malaria and sickle cell disease, once again for the two assumptions about the death rate for carriers discussed above. The figure shows that, at least within the range of the two estimates we have available, the assumption regarding the degree of protection afforded to carriers of the sickle cell mutation is not very important.

The fraction of the overall burden that takes the form of sickle cell disease rises with prevalence. For example, when $\pi = 0.20$, roughly one-tenth of the overall burden is in the form of deaths from sickle cell disease, with the other nine tenths due to malaria cases. When $\pi = 0.40$, sickle cell deaths account for roughly 20 percent of the burden. It is also of interest to calculate the net benefit of the sickle cell mutation, that is, the level of mortality in the presence of the mutation relative to the case where it is absent. The level of mortality absent the mutation can simply be read from the line representing M^{AA} , since in this case everyone would have the mortality rate of non-carriers. For example, in the case of 20 percent prevalence of the sickle cell trait, overall mortality due to malaria and sickle cell disease is 10 percent, but overall mortality due to malaria would be 11.1 percent if there were no sickle cell mutation. In areas of high malaria pressure, the value of the sickle cell mutation was higher. In the worst afflicted areas, the sickle cell trait reduced malaria related mortality by 20 percent.

We take sum of malaria and sickle cell deaths as our measure of the overall relative risk of AA vs. AS.

¹⁴Given a ratio of malaria mortality in the two groups, along with the equation for β , we can solve for the group rates of malaria mortality. These are $M^{AA} = \frac{1-\beta}{1-\beta X}$ and $M^{AS} = \frac{X(1-\beta)}{1-\beta X}$, where X is the ratio of M^{AS} to M^{AA}

mortality impact of the disease, which we call malaria burden.¹⁵ In Section 4 we explore the empirical relationship between the overall malaria burden and different measures of economic development at the ethnic group level, while in Section 5 we present a quantitative model for evaluating the expected impact of the malaria mortality burden.

3.4 Comparison of Malaria Burden to Modern Malaria Mortality Rates

For the WHO AFRO region, the under-five death rate from malaria is 0.59 percent per year. Multiplying this number by five gives an approximation to the probability of dying from malaria in the first five years of life, which is very close to the probability of dying from malaria before reproductive age.

The fraction of children who die of malaria is not exactly comparable to what we estimate in the historical data, because our measure M is the probability of dying of malaria conditional on not dying of something else. Such a measure is not usually examined in modern data, but it can be constructed relatively easily. Consider the case of Nigeria, which is a very heavily afflicted country. The life table for Nigeria for 2006 shows that the probability of woman surviving to age 25 is approximately 0.75. Taking this age to be the equivalent of adulthood in our historical data, we have:

$$0.75 = (1 - M)(1 - P) \quad (8)$$

Annual malaria deaths for children under five in Nigeria are estimated to be 8.8 per thousand, or 0.88 percent. This implies that roughly 4.4 percent of

¹⁵Our historical malaria burden measure is a potential alternative to the widely-used malaria ecology index of Kiszewski et al. (2004). That index takes into account both climatic factors and the dominant vector species to give an overall measure of how congenial the environment is to the spread of malaria. The index is calculated for grid squares of 0.5 degree longitude by 0.5 degree latitude. With the exception of New Guinea and some areas of southeast Asian, Africa is the only part of the world in which the index reaches its highest levels.

We compared the predictive power of these two measures for two different indicators of the malaria intensity in Africa (outside of Africa we do not view our measure as a viable alternative to the malaria ecology index because the sickle cell mutation is not present among many indigenous populations). The indicators were a dummy variable for having a highly malarious environment c.1900, based on Lysenko's (1968) map and average plasmodium falciparum transmission intensity in the year 2007 using data from Hay et al (2009). We used data at two different levels of aggregation: first, grid squares of 0.5 degree longitude by 0.5 degree latitude, and second, ethnic groups mapped by Murdock (1959). In all cases, the malaria burden measure had significantly greater explanatory power, as measured by R-squared by standardized regression coefficients, although when both measures are entered into a regression, they both remain significant (results available upon request).

An additional advantage of the malaria burden measure that is worth mentioning is that its quantitative interpretation is very simple: it represents the fraction of children expected to die from malaria or sickle cell disease, conditional on not dying of something else. By contrast, malaria ecology is an index of the stability of transmission of the malaria parasite and does not have a simple interpretation in terms of the human effect of the disease. Using it to assess the importance of malaria for health or economic outcomes requires scaling the malaria ecology index in a regression framework, as in Sachs (2003).

children will die of malaria before their fifth birthday. We assume that there are no further malaria deaths beyond this age. To incorporate this data into an estimate, we need to deal with the ambiguity of timing in the simple demographic model with which we started. Specifically, the model implies that a fraction $(1 - M)(1 - P)$ of children will survive both malaria and other conditions, but it is less clear about what those who don't survive die of.¹⁶ If malaria mortality comes before that from other conditions, then a fraction M will die of malaria and a fraction $(1 - M)P$ will die of other causes; if other conditions come first, then a fraction P will die of other causes and $(1 - P)M$ will die of malaria. The truth is obviously somewhere in the middle, and for lack of any firm data, we simply assume that deaths due to malaria and "other" had equal time profiles, which implies:

$$0.044 = M(1 - P) + \frac{PM^2}{M + P} \quad (9)$$

Solving for equations (8) and (9) yields $M=0.054$ and $P=0.207$. To compare the malaria burden, we can look at the fraction of the Nigerian population who are carriers of the sickle cell trait today. In the data of Piel et al (2010), 9.8 percent of adult alleles are of the S type (this is the average over the area Nigeria, weighted by current population density), implying that roughly 19.6 percent of adults are carriers.¹⁷ This in turn implies that slightly less than 10 percent of children would have died of malaria before reproductive age, conditional on not dying of something else. Thus the burden of malaria has fallen by a little less than half. Below we discuss the comparison of the level of decline in malaria mortality to the decline in mortality from other causes.

4 Assessing the Importance of Malaria to Early African Development

We turn now to an statistical analysis on the importance of malaria to early African development. Our data are at the level of ethnic groups. Analyzing data in this form, rather than countries or sub-national units, has two advantages. First, we expect ethnic groups to be less heterogeneous in terms of their levels of early development than political units, which are usually larger and were often created with little attention to realities on the ground. Second, the ethnic group data that we have allows us to get a picture of development further back in time than available data from political units.

In our first econometric exercise we focus on population density for 398

¹⁶The fraction who die is $M + P - MP$. This can be rewritten as $M(1 - P) + P(1 - M) + MP$, where the first term is children who died of malaria but would not have died otherwise, the second term is children who died of something else but would not have died of malaria, and the third term is children who died of one but would have died of the other.

¹⁷Luzatto (2012) presents very similar data for Nigeria, also derived ultimately from Piel et al.

ethnic groups from Alsan (2014).¹⁸ We also exploit a variety of information from Murdock (1967)’s Ethnographic Atlas, a database on 1,167 ethnic groups of six different regions of the world, including Sub-Saharan Africa. This database includes information from the level of subsistence economy to the degree of political integration of each ethnic group.¹⁹ In order to account for the potential confounding effects of a range of geographical and climatic factors, we also construct different geographic and climatic variables by combining georeferenced data with a digitalized version of Murdock (1959)’s ethnic maps.

In order to assess the effect of malaria environment in our measures of early development we estimate different specifications of the following equation ²⁰:

$$y_i = \alpha + \beta M_i + \gamma' G_i + \delta' W_i + \rho' T_i + \lambda' X_i + \mu_C + \epsilon_i \quad (10)$$

Where the subscript i denotes ethnic group, y is a measure of early development, and M is our measure of historical malaria burden. The vectors G , W , T , and X include geographic, access to waterways, climatic, and other plausibly exogenous controls, respectively. μ_C is a collection of 41 ethnographic cultural cluster dummies from Murdock (1967). We describe these controls in detail below. Finally, ϵ is the error term that is allowed to be heteroskedastic.

4.1 OLS Estimates

The effect of Malaria on Population Density Circa 1900.

Table 2 presents the first statistical results where population density is our dependent variable. We focus on population density following the usual Malthusian reasoning that, holding land quality constant, steady state population density is a good indicator of the state of development. In order to ease the interpretation of the results and compare them across the different specifications, we report standardized coefficients.²¹

In column 1 we only control for a set of geographic controls including land quality, elevation, and ecological diversity. Land quality is the mean value of the index of land suitability for agriculture from Ramankutty et al. (2002) and represents the probability that a particular grid cell (of the size of about 50 x 50 km) may be cultivated. To account to the fact that malaria is more prevalent in less ecologically diverse areas which also tend to be less densely populated, we

¹⁸We thank Marcella Alsan for providing population density data. Alsan (2014) exploits population figures and total land area for each ethnic group -as implied from shape-file from the digitalized version of the ethnicities’ map- in Murdock (1959).

¹⁹Since there is no perfect match between ethnicity names in Murdock (1967)’s Ethnographic Atlas and Murdock (1959)’s map we follow Fenske (2013) to match a total of 520 ethnic groups.

²⁰We use our malaria burden measure for the case in which AS relative malaria mortality of 0.08 in all the empirical analysis in this paper. Using a different assumptions regarding the relative protection of AS genotypes does not affect our statistical results.

²¹The standardized coefficients report the number of standard deviation changes in the dependent variable for a one-standard deviation change in the independent variable.

include a measure of ecological diversity from Fenske (2013).²² Finally, we also include mean elevation of terrain to account for the fact that elevation has a strong negative effect on malaria prevalence and may also have an independent effect on development.²³ Interestingly, the coefficient estimate for β suggests a strong *positive* statistically significant relationship between our measure of malaria burden and population density. This striking positive association is also documented in other works based on modern data at the country level (such as in Gallup, Sachs, and Mellinger 1999).

The presence of water bodies may positively correlate with both population density and malaria prevalence (such as around Lake Victoria). In column 2 we add the vector W which includes the log total area of water bodies accessible for the ethnic group (note also that humidity tends to be high near rivers facilitating mosquito breeding) and the log of the geodesic distance of the centroid of the historical homeland of each ethnic group from the nearest coastline (the greater the distance to the coast, the lower the population density). Nonetheless, the point estimate remains virtually unaltered (column 2).

We next include a set of climatic controls (column 3). We pay special attention to temperature which plays a key role in determining the suitability for the transmission of malaria (Gething et al, 2011). Using monthly temperature data from the CRU TS3.21 dataset we calculate mean temperature over the time period 1901 - 2012 for each high-resolution grid (0.5 x 0.5 degree). We then compute mean and standard deviation for each ethnic group homeland in Murdock (1959). The introduction of these controls does not affect the previous results. This is not surprising since malaria's temperature dependence is highly non-linear. Therefore, including temperature in levels is likely to be only indirectly linked to the intensity of malaria transmission (Gething et al 2011). In column 4 we add a set of 41 cultural province dummies. Although the estimated coefficient on malaria is reduced somewhat, it remains positive and strongly statistically significant.

Another potential bias in the estimation of β results from the omission of ethnic-group agricultural practices. Using data for 60 communities in East and West Africa, Wiesenfeld (1967) shows a positive correlation between the prevalence of the sickle cell trait and the cultivation of crops associated with swidden agriculture. Swidden agriculture (clearing small areas of forest with slash and burn practices) multiplies the number of breeding places for *Anopheles Gambiae*. Roots and tree fruits are the main crops associated with this agricultural method and the most important ones for the African case are yams and bananas. To account for this potential omitted variable, we add the mean suitability for cultivation of yams and bananas in column 5.²⁴ There is little effect on our

²²This measure of ecological diversity consists in a Herfindhal index based on the territorial shares of different ecological types within the boundaries of each ethnic group's homeland

²³For instance, Gallup, Sachs, and Mellinger (1999) shows that population density is greater (lower) at high altitudes in the tropics (temperate zones).

²⁴We use suitability for cultivation of yams and bananas rather than data on actual cultivation of these crops (which is available in the Ethnographic Atlas) because the latter is likely to be endogenous. Specifically, Wiesenfeld (1967) argues that the sickle-cell trait could

estimate of β . To account for the potential confounding effect of the tsetse fly, we also include Alsan (2014)’s tsetse suitability index in column 6. The point estimate remains unaltered. In sum, after controlling for a battery of geographic, waterways, climates, and regional controls, we find that one standard deviation increase in our malaria burden measure is statistically associated with an increase of almost one-fifth of a standard deviation of population density.

Finally we check that this striking positive association documented so far is not driven by our measure of historical malaria burden. In column 7 we replace our measure by an alternative proxy for the historical severity of malaria. This alternative measure is constructed based on Lysenko’s (1968) map categorizing the world at the beginning of the twentieth century into six classes of malaria transmission intensity in order of severity: no malaria, epidemic, hypoendemic, mesoendemic, hyperendemic, and holoendemic. We focus in the two worst malaria environments as highly malarious, and construct a measure which is the fraction of the territory of our unit of analysis in which malaria fell in this range. We again find a positive and strongly significant association between malaria severity and population density.²⁵

Other Measures of Development.

Before turning to an instrumental variable approach we repeat the previous exercise to examine the relationship between malaria burden on other measures of ethnic development. Following Nunn and Wantchekon (2011), we exploit a rich source of ethnic information provided by the Murdock (1967) and estimate a new version of equation (10) where our dependent variable is now a given measure of ethnic development. Table 3 presents the main results. All the specifications include the full set of controls as in column 6 of Table 2. Since we no longer rely on the availability of population figures, the sample size for each specification will depend only on the availability of the ethnographic variable for ethnic groups that can be located in the Murdock’s map (our largest sample consists of 520 ethnic groups).

All the dependent variables in Table 3 are binary, with 1 indicating a higher prosperity level (see note in table) and come from Ethnographic Atlas, except for the variable in column 1 which is constructed using Bairoch (1988).²⁶ We first study the statistical relationship between malaria burden and the probability of at least one city with population above 20 thousand people located in the homeland of the ethnic group by the year 1850. We find no statistically significant association between this measure of urbanization and historical malaria burden. In column 2, we find a statistically significant and positive association between our measure of malaria burden and the complexity of settlement pattern of the ethnic group (i.e: whether the ethnic group settlement pattern is

reduced the environmental limitation of malaria, allowing populations to embrace slash-and-burn practices that have high yields per unit labor.

²⁵We also experimented with other specifications including region dummies (i.e: West, East, Central, North, and South) and country dummies. The results are not substantially affected.

²⁶All the results in Table 3 are OLS coefficients. We also estimated probit and ordered probit (for the original categorical variables) and reached qualitatively similar results.

either compact and relatively permanent or complex, as opposed of being fully nomadic or semi sedentary). These results suggest that malaria did not impose important impediments to the creation of complex settlement and urbanization.

We next look at the statistical association between our malaria burden and precolonial political institutions. Following Gennaioli and Rainer (2007), we construct a measure of centralization of political power that takes the value 1 if the ethnic group has any jurisdictional level transcending the local community. In line with previous statistical associations, point estimate from column 3 suggests that ethnic groups located in areas with worst malaria burden do not tend to have less sophisticated political institutions beyond the local communities. We next investigate whether malaria burden negatively impacts the ability of an ethnic groups to generate agricultural surpluses. From column 4 in Table 3, malaria burden is not statistically associated with intensive agriculture being the main contributor of the subsistence economy (as opposed to gathering, fishing, hunting, pastoralism or casual and extensive agriculture). In sum, all the standardized coefficients reported in Table 3 suggest an statistically weak and economically small association between our malaria burden measure and different measures of pre-colonial prosperity.

4.2 Instrumental Variable Approach

We view the previous analysis as exploratory. In the absence of a source of exogenous variation on historical malaria burden, it is difficult to identify the precise impact of malaria on development in the past with any reasonable degree of certainty. Although we attempt to account for several confounding factors in our analysis, bias due to measurement error in our malaria burden variable and reverse causality is likely to be present in the point estimates we report above. In this section we exploit variation in an temperature-dependent index of suitability for transmission of malaria developed by Gething et al (2011) to instrument for our historical malaria measure. Temperature affects the life cycles of both the plasmodium parasites and mosquito vector, and thus it shapes the landscape of malaria.

Gething et al (2011) present a dynamic biological model incorporating the key mechanisms through which temperature affects malaria transmission to create high resolution maps of temperature suitability for malaria transmission. Unlike previous attempts to map malaria environmental suitability, their measure incorporates the effects of continuously changing temperature regimes on vector and parasite populations. It is also important to note that this measure does not depend on the actual distribution of mosquito vectors, which can be influenced by human economic activity (see, for example, McCann 2011). In this way it avoids a potential endogeneity problem with Kiszewski et al. (2004) malaria ecology measure.

First Stage and Comparison to Malaria Ecology

Gething et al (2011) provide two suitability indices, one for plasmodium falciparum and another for plasmodium vivax. The two are highly correlated, so we simply take their average, which we refer to as plasmodium suitability. In Table 4 we look at the ability of plasmodium suitability to predict malaria burden. All the specifications account for the potential direct impact of altitude, proximity to water, agricultural suitability, ecological diversity, and cultural cluster dummies.

We first show that plasmodium suitability is a strong linear predictor of malaria burden even when the first and second moments of historical temperatures are accounted for. This is an important check, since our proposed instrument is constructed from temperature data, and temperature can indirectly impact development through agricultural productivity or other diseases. Therefore, after controlling for the mean and standard deviation of temperature, the variation in the instrument that we are actually exploiting comes from non-linear effects of the time pattern of temperature variation on mosquito survival. The highly statistically significant point estimate in column 1 suggests that one standard deviation increase in the plasmodium suitability index predicts almost half of a standard deviation increase malaria burden. As expected, accounting for the climate controls in column 2 reduces the size of the point, but the standardized beta coefficient of 0.41 remains highly significant.

We next run the same exercise using malaria ecology as main explanatory variable for malaria burden. When none of the climate controls is added, column 3 in Table 4, malaria ecology is a strong linear predictor of malaria burden. However, its predictive power as measured by the standardized coefficient is one fifth the size of coefficient for the plasmodium suitability index. When the climate controls are added to the specification in column 4, we find no statistically significant association between malaria ecology and malaria burden.

Finally, we run horse races between our proposed instrument and malaria ecology and find that, regardless of the addition of climate controls, the plasmodium suitability index outperforms malaria ecology as a predictor for malaria burden.²⁷

Instrumental Variable Estimates

Table 5 presents IV estimates for our preferred specifications from Tables 2 and 3 (that is, including all the aforementioned controls). For the specifications where the outcome variables are either population density or indicators for the existence of large city, centralized political power, and intensity of agriculture, the signs of the point estimates reverse, but they remain statistically insignificant.

²⁷To test whether this same pattern hold using a more recent measure of malaria's impact, we ran similar regressions using the average plasmodium falciparum transmission intensity (parasite rate) in 2007 from Hay et al. (2009) as a dependent variable. The results were very similar to those in Table 4: the plasmodium suitability index outperforms malaria ecology as a linear predictor of malaria endemicity, and the predictive power of malaria ecology is statistically weaker when controlling for the mean and standard deviation of temperature.

Therefore, even after instrumenting our measure of historical malaria burden with plausible exogenous variation in malaria suitability we do not find strong evidence that the impact of malaria would have been very significant on early African development. Taking the case of population density as an example, the IV point estimate - albeit highly statistically insignificant- suggests that one standard deviation increase in malaria burden is associated with a small effect on population density: a reduction of only six percent of its standard deviation.

In Table 6 we provide additional robustness checks. Each cell in Table 6 reports the coefficient on the historical malaria burden from a separate IV regression depending on the outcome and the specification we ran. Below the standard errors, the first stage F-statistics are reported in brackets. The outcome variable of interest is listed at the left, while the information on the econometric specification is noted at the bottom.

Some of our early development indicators may be influenced by external factors. If some ethnic groups located in places where the disease environment and other geographic factors handicapped development were also the most affected by the colonial rule, then not taking into account cross-ethnicity variation in colonial power might lead to an important bias in the estimates of the effect of malaria on development. For the specifications in column 1 we add three variables accounting for external influence in the development of the ethnic group. In particular, we account for the importance of slave trades for the period 1400-1900 (i.e: log of 1+ total slaves exports normalized by area of homeland), and two dummy variables indicating if European explorers traveled (between 1768 and 1894), or a historical trade route passed through, the homeland of the ethnic group. Although we acknowledge a potential endogeneity of these controls, their addition does not seem to affect previous results.

Studies in molecular genetics found that the S gene is associated with at least four different beta-globin haplotypes in Africa, which differ in at least three different genetic markers, providing evidence of the multicentric origin of the mutation. We construct the distance from each ethnic homeland to the closest hypothesized origin of the mutation. Although this measure is strongly correlated with malaria burden, its inclusion as a control in column 2 does not affect the previous results. In column 3 we include the prevalence of the Hemoglobin C mutation as a control and we still find that malaria burden does not impose a negative effect on early development. For the specifications in column 4 we add all the previous controls together and we find no change in our previous results.

For the specifications in column 5 we only consider ethnic groups located in tropical Africa. Again we do not find statistical evidence that malaria burden negatively affected population density, settlement complexity, centralization of power, or the intensity of agriculture. We do find, however, that for ethnic groups located in tropical Africa, one standard deviation increase in our measure of historical malaria burden is related to a 12 percent reduction in the probability of having a large city in 1850 (significant at the 10% level).

To summarize, exploiting plausible exogenous variation from a temperature-dependent index of suitability for transmission of malaria we do not find evidence

that the impact of malaria would have been very significant on early African development. We do not find any negative statistically significant association between our measure of malaria burden and population density when we focus on ethnic groups. We do not find consistent evidence of any negative effect on other measures of ethnic development and prosperity such as settlement patterns, the degree of subsistence economy, and one measure of sophistication of the political institutions.

5 Model-Based Estimates of the Economic Burden of Malaria

5.1 Direct Effect of Malaria Mortality

Current analyses of the burden of disease focus on measures such as years of life lost or disability free life years lost. From this perspective, the death of a young child is particularly costly because he or she had so many potential life years. The ethical considerations regarding the allocation of scarce lifesaving resources, and implicitly the cost of death and disease experienced at different ages, are quite complex (see Persad, Werthheimer, and Emanuel, 2009). In assessing the role that disease played in affecting development historically, however, it seems reasonable to take a purely instrumental view of life and health, in which the primary considerations are how much society has invested in an individual and that individual's potential to produce services for society in the future.

To formalize this idea, consider a simple model of production and consumption with individuals of different ages. Let c_i be the consumption of an individual of age i . We assume that there is no storage of output between periods. Let N_i be the number of people in age group i . The social budget constraint is

$$Y = \sum_0^T N_i c_i \quad (11)$$

Where Y is total output and T is the maximum lifespan. We assume that consumption at each age is determined by two things: a consumption level of individuals at some benchmark age (for example, prime age adults), which we call $\bar{c}$, and some age-varying relative consumption coefficient $\tilde{c}$.

$$c_i = \tilde{c}_i \bar{c} \quad (12)$$

The values of $\tilde{c}$ presumably reflect changing biological needs for consumption over the course of the life cycle as well as the arrangements by which consumption is divided up among different groups in society. One would not necessarily expect the pattern of consumption to be the same in all societies at all times. However, as discussed below, available data do not vary all that much.

We assume that output is produced by a fixed quantity of land, X , and labor. Let $\tilde{l}_i$ be the quantity of labor supplied by an individual of age i , again, relative to some benchmark level. As with consumption, the pattern of relative

labor input across age groups reflects both biological differences and differences among societies in the economic value of different characteristics (for example, strength vs. wisdom vs. manual dexterity). Taking the production function to be Cobb-Douglas, we have

$$Y = AX^\alpha \left(\sum_{i=0}^T N_i \tilde{\ell}_i \right)^{1-\alpha} \quad (13)$$

Letting N be the total population and $\tilde{N}_i$ the fraction of the population in age group i , we can solve for the level of benchmark consumption as a function of population size and demographic structure:

$$\bar{c} = AX^\alpha \left(\frac{X}{N \sum_{i=0}^T \tilde{N}_i \tilde{\ell}_i} \right)^\alpha \left(\frac{\sum_{i=0}^T \tilde{N}_i \tilde{\ell}_i}{\sum_{i=0}^T \tilde{N}_i \tilde{c}_i} \right) \quad (14)$$

The first term in parenthesis is the ratio of land to efficiency-adjusted labor. It affects the level of benchmark consumption for the usual Malthusian reasons. The second term in parentheses can be thought of as the demographic efficiency of the economy. Consumption will be higher in a society where the population is distributed relatively more toward ages in which people supply labor relative to ages in which they consume.

We now turn to the determination of population density. Our assumption is that in the historical period we are examining, Africa was in Malthusian population equilibrium. In such an equilibrium, as long as the quantity of land is constant and the production technology is fixed, the growth rate of population is zero and the population size is constant (or, more realistically, the population size and standard of living are mean reverting in response to stochastic shocks). The near constancy of human populations over long historical eras suggests that such a mechanism must have been operating in most places for most of human history (Galor and Weil, 2000). Although there are clearly times when Malthusian constraints were lifted (such as the peopling of the Americas in pre-historic times), a simple analysis of the effect of compounding shows that most of the time, average population growth rates must have been quite close to zero. Without a homeostatic model of the Malthusian sort, it is vanishingly unlikely that such near constancy of population would obtain.

Herbst (2000) argues that the abundant land was a persistent characteristic of African economies, as a result of which it was the control of people, rather than territory, that was of primary interest to rulers. In our view, however, historians are too quick to dismiss the Malthusian model without having another explanation for the near constancy of population over time. It is conceivable that, rather than land being the binding resource, the role of equilibrating population size was played by density-dependent disease. This is the view taken by McGuire and Coelho (2011). One can work through a model with density

dependent disease producing results very similar to those we show here.²⁸

Following Kremer (1993) and many subsequent papers, we model the Malthusian equilibrium by assuming that population size adjusts such that the level of consumption (for the benchmark age) is held constant. In the empirical sections above, the measure of development on which we focus is population density. Here, rather than looking at the simple population density (that is N/X) we do a demographic adjustment. The idea is that, just as a society populated by a high number of working age adults relative to children can produce more output per capita, a society populated by a high number of working age adults relative to children is also more “dense” for a given level of population size. We thus scale the number of people by their labor input weights in defining effective density. That is

$$density = \frac{N \sum_{i=0}^T \tilde{N}_i \tilde{c}_i}{X} = \left(\frac{A}{\bar{c}} \right)^\alpha \left(\frac{\sum_{i=0}^T \tilde{N}_i \tilde{c}_i}{\sum_{i=0}^T \tilde{N}_i \tilde{c}_i} \right)^{1/\alpha} \quad (15)$$

For our primary exercise, we consider changes in age structure and density that hold the benchmark level of consumption $\bar{c}$ constant

To use this equation to assess the effect of disease mortality on the consumption benchmark, we need estimates of the consumption and labor input profiles and the exponent on land in the production function, as well as an estimate of the effect of disease on the age structure of the population.

Consumption and Income Profiles

A number of sources provide data on the life cycle profiles of consumption and labor input. Mueller (1976) synthesizes data from nine societies practicing what she calls “peasant agriculture”, by which she means agricultural systems

²⁸Several recent studies have argued that higher disease, through its effect on the ratio of population to land, could raise income in some circumstances and conversely that health improvements could lower income. Voigtlander and Voth (2013) argue that in Europe before the industrial revolution, income rose as a result of increased mortality due to plague, urbanization, and warfare, and that this rise in income was instrumental in knocking the continent onto the path of industrialization. Acemoglu and Johnson (2009), in their analysis of the worldwide epidemiological transition of the mid 20th century, find that countries that experienced greater increases in life expectancy saw slower growth in income per capita. And Young (2005) claims that higher mortality and lower fertility due to HIV in South Africa will more than compensate for declines in worker productivity due to the disease, so that income per capita will actually rise as a result of the epidemic. Ashraf, Lester, and Weil (2008) examine the Malthusian channel as part of a broader analysis of how health improvements affect growth. They find that higher survival indeed pushes income lower through the Malthusian channel (as well as through capital dilution), although this effect is more than compensated for in the long run by higher labor productivity due to better health. Such an effect could be induced in our model as well, if increases in the benchmark level of fertility, $\bar{c}$, are required to increase fertility in order to make up for higher mortality. This would be an additional channel (beyond the demographic efficiency that we examine) by which higher mortality would lower population density.

which use primarily traditional methods of cultivation, small landholdings, and low capital inputs. South and Southeast Asia are primarily what she has in mind. The profiles are shown in Figure 4. Note that Mueller's data on labor input apply to production of output as it would appear in measured GDP but exclude home production. Much of the latter is done by women, so in her data, productivity by prime age women is only 30% of the level of similarly aged males. Because we are interested in both sorts of output, we focus on the male profiles. For consumption, Mueller provides two profiles in which consumption of people at each age (and of each gender) is compared to males aged 20-54. For the medium consumption profile, children age 0-4 have a value of 0.32; for the low consumption profile, the value is 0.12 (prime aged women get a value of 0.80). We use the medium profile.²⁹

Figure 4, along with the life table examined below, can be used to think about the cost to society of deaths at different ages. The most costly deaths are those that take place in young adulthood, at the age where the consumption and income curves cross. These are individuals in whom society has invested appreciable resources (food, childcare, education), and who have ahead of them many expected years producing a surplus. Deaths of people at older ages are not as costly, since such individuals have fewer expected years of surplus production; and similarly deaths of very young children are less costly because society would still have to invest in them for many years before they started producing a surplus. Under this interpretation, malaria deaths are relatively low cost, with the exception of deaths of women in their first pregnancies, who are near the peak of their value as assets to society in terms of the balance between resources invested in them and services they can deliver.

Land in the Production Function

A number of studies have modeled pre-modern Malthusian economies and/or the agricultural sectors of more modern economies, in which the only inputs into production are land and labor. Kremer (1993) uses 1/3 as an upper-end estimate of land's share, based on evidence from sharecropping contracts. In the model of Hansen and Prescott (2002), the Malthus sector, which is the only part of the economy producing output prior to the industrial revolution, has a land share of 0.3. Stokey (2001) applies a Cobb-Douglas production function to the agricultural sector for Britain in 1850, with an exponent on land of 0.45. In the calculation reported below we use a value of one-third as the land share in

²⁹We also examined data from two other sources, which paint a similar picture. First, Lee and Mason (2010) look at data from four contemporary hunter-gatherer societies. The underlying data are in terms of calories collected and consumed. Their data are quite similar to the profiles in Mueller. Second, Lee and Mason also look at data for the four poorest countries that are part of the National Transfer Account project: Kenya, Philippines, Indonesia, and India. Income in this case is labor income, including unpaid family labor, and pertains to both men and women. These data differ from the other two sources primarily in showing a decline in income of the elderly that is not present in the peasant agriculture or hunter-gatherer data.

agriculture.³⁰

Population Age Structure

The other piece of data used in the equation (15) is $\tilde{N}_i$, the relative number of people in each age group. In general, this will be a function of both the probability of survival to each age and the history of births or population growth. However, for the long historical periods that we are considering, population growth must have been very close to zero, which implies a constant number of births per year. We can thus approximate the age structure of the population $\tilde{N}_i$ with the fraction of survivors at each age from the life table. Thus our approach to assessing the role of disease in affecting consumption possibilities is to start with a baseline life table and then consider alternations that would result from eliminating or adding particular sources of mortality, using information on the age of death from different diseases.

This exercise requires choosing a baseline life table that represents Africa in the pre-colonial period. Systematic data on life expectancy in Africa is widely available only starting in the 1950s. For the period 1950-55, the United Nations estimate of life expectancy at birth in sub-Saharan Africa is 37.8 years (United Nations, 2009). Acemoglu and Johnson (2007) date the beginning of the “international epidemiological transition,” driven by more effective public health measures, the discovery of new chemicals and drugs, and international interventions to 1940. Although the transition came late to Africa, it is very likely that the 1950-55 figure represents an improvement relative to previous decades. Clearly after 1955 the pace of change was rapid. The UN estimates that life expectancy in sub-Saharan Africa rose by two years in each of the subsequent five year periods. Further evidence that health improvements were already underway by 1950 comes from data on total population size. Africa’s population grew at a rate of 1.0 percent per year between 1900 and 1950, compared to a growth rate of 0.2 percent per year over the previous century (United Nations, 1999).

Data for the period prior to 1950 are very sparse. Acemoglu and Johnson (2007, Appendix C) pull together disparate sources to present a few estimates for the period before 1950. These are, Angola in 1940: 35 (both sexes); Mozambique in 1940: 45 (both sexes); Ghana in 1948: 38 (both sexes); and Mauritius, 1942-46: 32.3 (male) and 33.8 (female). Riley (2005) estimates that prior to the health transition (he uses a different definition than Acemoglu and Johnson) that began in Africa in the 1920s, life expectancy at birth averaged 26.4 years (this is the mean of 12 estimates, which range from 22.5 to 31).³¹ In Asia, life

³⁰The other issue relevant here is the elasticity of substitution between land and other inputs. The studies discussed here all use a Cobb Douglas production function, in which the elasticity of substitution is one. Weil and Wilde (2009) use modern data on the share of natural resources paid to fixed factors of production to estimate this elasticity of substitution. Their estimate is close to 2, although very imprecise. The higher is the elasticity of substitution, the larger will be the effect of change in demographic structure on population density.

³¹Some representative values are Angola: 27 years in 1940; Egypt: 30-33 years in the 1930s; Ghana: 28 years in 1921; Kenya: 23.9 years in the 1930s; South African black population:

expectancy prior to the health transition, which started there between 1870 and 1890, was 27.4. In Europe the transition started in the 1770s, and prior to it life expectancy was 34.3.

Riley comments that available estimates of African mortality prior to the health transition all come from European colonies in Africa. There is a reasonable basis for thinking that life expectancy may have been higher prior to colonization, the arrival of Arabic speaking merchants, and the dislocations produced by the slave trade. Unfortunately, almost no information for this period is available. Steyn (2003) examines mortality in the pre-colonial period in northern South Africa through an examination of skeletal remains. She estimates life expectancy in the period 1000-1300 AD at 23.2, with the probability of surviving to age 20 being 48 percent. Remains for the post-1830 period show a slight decline in life expectancy after the expansion of European influence.

Based on the above discussion, we use as our baseline the United Nations (1982) model life table for a population with life expectancy at birth of 35 years (male and female combined).³² This is the lowest life expectancy for which the model life table is available.³³

To apply this framework to different diseases, one needs a profile of the mortality effect of the disease at different ages. For malaria, we use age-group death rates from Murray and Lopez (1996).³⁴ To model the effect of malaria on the historical life table we proceed as follows. We start with the UN life table (life expectancy of 35) discussed above. We take this as a benchmark. To this we add additional mortality at each age proportional to the current age profile of malaria deaths. Specifically, we take the current death rates from malaria and multiply them by a scaling factor, and add these death rates to the death rates in the UN life table. The scaling factor is chosen to match the magnitude of malaria deaths we want to model. Above, for the case where sickle cell prevalence is moderately high ($\pi = 0.20$), we showed that M , the implied probability of dying from malaria conditional on not dying of something else prior to childbearing,

38.1-40 years in 1935-40; Tunisia: 28.8 in the 1920s; Uganda: 23.9 in the 1930s; Zimbabwe: 26.4 in mid 1930s.

³²Unfortunately, the model life tables that are available, to the extent that they reflect African data at all, certainly do not reflect the pattern of age-dependent mortality that existed in the period before the modern health environment of both disease and treatment was in place. It is likely that the pattern of mortality, and in particular the ratio of deaths at different ages, differed from what is observed today, but a priori there is no basis for guess the nature of this difference.

³³Once we have a life table, we can also address the issue of how the decline in the non-malaria death probability P compares to the decline in the malaria death probability M . The UN Model life table with life expectancy of 35 (general model, for females) implies 57.2 percent of girls will survive to age 25, which we can take as an average for childbearing. Thus, $0.572 = (1 - M)(1 - P)$. Using a value of $M = 0.10$ (consistent with the prevalence of the S allele in Nigeria, as discussed above) implies a value of $P = 0.364$. As discussed above, our analysis of modern data in Nigeria imply $P = 0.207$ and $M = 0.054$. Thus the proportional decline in malaria mortality has been almost exactly the same as the decline in mortality from other causes. However, this calculation is very sensitive to the choice of life expectancy in the historical period.

³⁴These are 0.00559 for ages 0-4; 0.00042 for ages 5-14; 0.00033 for ages 15-44; and 0.00036 for ages 45-59.

was 0.10. We set the scaling factor to match this probability, taking 25 to be the age of childbearing.³⁵ The implied scaling factor is 2.95. Figure 5 shows the survival curves for the baseline case and the case with high malaria.³⁶ Life expectancy falls by 5.1 years from this additional malaria mortality.

Effects of Malaria on Density

Applying equation (15), we can calculate population density using both the original life table and the life table modified to reflect malaria mortality. Our finding is that adding malaria mortality reduces density by only 2.8 percent. This is obviously an extremely small effect for such a large change in mortality. We can put further this number in context two different ways. First, Alsan (2014) also examining African data, finds that a one standard deviation change in an index of tsetse fly impact leads to a 45 percent change in population density in the period prior to European settlement. This effect is much larger than ours. Second, we can compare the effect of malaria to variation in land quality. Ashraf and Galor (2011), using data for 1500, estimate the coefficient in a regression of log population density on the log of a land quality index as 0.587.³⁷ A reduction in density of the same magnitude as the one we attribute to malaria could thus be induced by a reduction of 0.048 in the log of land quality, which is less than one twenty-fifth of the standard deviation of that measure, once again showing that the effect of malaria on density that we estimate is small.

The reason that our estimate of the effect of malaria is so small is two-fold. First, malaria deaths are concentrated at young ages, and second, consumption of young children is low relative to consumption of adults. Putting these together, most deaths from malaria do not, in this model, represent a significant loss of resources to society. In our calculation, deaths beyond age five account for only 1/3 of the reduction in life expectancy due to malaria, but for 2/3 of the economic cost of the disease.

We can easily compare the implied effect of malaria on density in this calculation with the regression coefficients estimated in the previous section. The calculation here says that going from a malaria burden of zero to a malaria burden of 0.1 (which is consistent with equilibrium prevalence of the sickle cell trait $\pi = 0.2$) would reduce density by 2.8 percent. In a regression of log population density on the level of malaria burden, this variation would produce a regression coefficient of -0.28. This is very close to zero in comparison to the coefficient on malaria burden in regression with log population density as the dependent variable (the first column of Table 5, but in non-standardized form), which is -2.38.

³⁵Specifically, we choose the scaling factor x such that: $0.90 = (1 - 0.00559x)^5(1 - 0.00042x)^{10}(1 - 0.00033x)^{10}$

³⁶Of course the UN table already includes deaths due to malaria. Another approach would be to take as the baseline a current life table stripped of malaria mortality, and then to add malaria mortality back in. This procedure would yield very similar results.

³⁷Table 2, column 2; the t-statistic is 8.3.

We assess the robustness of this result with respect to three variations on the calculation. First, we consider the cost of gestation. In the model presented above, the cost of death of a newborn is zero, which is clearly unrealistic. Having said this, we do not have a good estimate for the consumption price of producing a newborn child. For lack of better data, we assume that the price is equivalent to one year's consumption for a child aged 0-4, which in Mueller's data is 0.32 of the consumption of a prime age adult. Implementing this change increases the reduction in density from 2.8 percent to 3.4 percent.

The second variation that we consider is to allow for costs from maternal mortality. The death of a child necessitates, in equilibrium, an additional birth, which in turn exposes a prime age woman to mortality risk. Maternal mortality is far more costly to society than the death of a child, because a prime aged woman has already passed through the years in which her consumption was higher than her labor input: using our model, the cost to society of a 1 percent chance of death of a one-year-old is only 0.011 percent of average consumption, while the cost of a 1 percent chance of death of a 25 year old is 0.14 percent of average consumption. To see how this effect raises the economic burden of malaria, we would ideally want an estimate of the maternal mortality rate (MMR) in Africa during the pre-colonial period, which is not available. As an alternative approach, we ask how large the MMR would have to be so that the burden of malaria we calculate was doubled. The answer is that this would require an MMR of 7 percent. To benchmark this number, MMRs in the worst-afflicted developing countries today are roughly 1 percent, while in historical data from Sweden, the MMR in 1750 was 0.9 percent (Hagberg and Wall, 1986). It thus seems to us unlikely that this channel would have a major quantitative impact on our estimate of the economic burden of malaria.

The third variation that we consider is lowering the baseline life expectancy in our calculation. Intuitively, the death of a child is costly to society because of the resources that have to be used to replace him or her. The higher is baseline mortality, the more expensive it is to produce a child of given age, because there is a lower chance of a newborn reaching that age. As mentioned above, the lowest life expectancy in the UN model life tables is 35 years. However, using the Brass life table methodology, we can generate a family of life tables with the same general shape as the UN table, but with lower life expectancies.³⁸ We do this to create life tables with life expectancies at birth of 30, 25, and 20 years. We then adjusted these life tables with the addition of age-specific malaria mortality, as described in our baseline analysis. The result is that in the presence of malaria, population density is reduced by 3.5, 4.6, and 6.4 percent, respectively. Thus it is true that in the presence of lower life expectancy, the impact of malaria on population density is amplified, but for reasonable values

³⁸The Brass transformation is as follows. Let $l(x)$ be the probability of living to age x in a given life table (we use the UN life table with life expectancy of 35). We can generate an alternative table via the transformation $l^* = (1 + \exp(2\alpha + 2\beta\lambda(l(x))))^{-1}$, where $\lambda(l(x)) = \ln \left[\frac{1-l(x)}{l(x)} \right] / 2$. Holding the value of β constant at one, we vary the value of α in order to change life expectancy while maintaining the general shape of the survival curve. See United Nations (1983), chapter 1.

of life expectancy the effect remains small. (Note that when life expectancy in the absence of malaria is 20 years, life expectancy inclusive of malaria is 16.0 years, which is significantly lower than any historical estimate that we know of)

The above analysis shows that, at least in terms of the extra costs associated with raising children who were subsequently going to die of malaria, the effect of the disease on population density in Africa should have been relatively low. However, it is possible that there are other channels through which high malaria mortality may have mattered. Death from disease may impose an obstacle to economic development beyond the instrumental cost discussed here. Stone (1977) argues that in the face of high child mortality, pre-modern Europeans avoided forming emotional bonds with their young children, even to the point of giving the same name to two living children, expecting only one to survive. In such an environment, one might also expect to see less investment of tangible resources in the human capital (both health and education) of their children. Thus even if high child mortality from malaria were not important in the strict sense of draining resources from the economy, it might have contributed to reduced investment in children. Several recent theories of economic growth have stressed the importance of such investments to long-term development.

Another possibility is that malaria affected development in Africa through morbidity (ill health) rather than mortality. This in turn could be via two channels: individuals suffering episodes of malaria infection were less productive, or episodes of malaria infection early in life produced lasting decrements to individual productivity.³⁹ Ashraf, Lester, and Weil (2008) evaluate both of these channels in the context of modern Africa, finding their economic impact to be small. For example, they calculate that the long-run effect of a complete eradication of malaria in Zambia, which has a relatively typical burden, would result in labor input per worker rising by 1.5% in the long run (this includes the effect of induced increases in schooling). Part of the reason for this relatively small effect is that malaria morbidity is simply not as important as morbidity from other diseases, and this in turn is due to the short average duration of malaria episodes among adult. Malaria accounts for only 5 percent of total life-years lost to disability, with most of this resulting from cases during the first five years of life (Murray, 1996). Unfortunately, no data are available regarding how malaria morbidity historically compared to today, and there is no genetic evidence that can be brought to bear as is the case with malaria mortality.

³⁹In the context of a modern economy, a potentially important channel by which malaria affects economic outcomes is through schooling. Bleakley (2010), Cutler et al. (2010), and Lucas (2010) all use national anti-malaria campaigns in the middle of the 20th century as quasi-experiments in order to study the effect of childhood exposure to the disease on human capital accumulation and adult economic outcomes. Their findings are highly variable, with Cutler et al. estimating a small effect and Bleakley a very large one. Lucas, whose findings fall between the other two, estimates that a 10 percent reduction in malaria incidence raises completed schooling by 0.1 years.

6 Conclusion

Malaney, Spielman, and Sachs (2004) write, “how better to evince the power of the parasite than with a potentially lethal modification of the genetic code as a desperate Darwinian defense against the even more deadly ravages of malaria? Accordingly, it may be expected that a force strong enough to rewrite our DNA will rewrite many of the lives and economies that it touches.” In this paper we have tried to address this issue directly. That is, we have used the extent which malaria left its mark on the human genome to back out the severity of the disease’s impact, and then in turn we have tried to assess how large the economic impact of that disease would have been.

In areas of high malaria transmission, 20 percent of the population carry the sickle cell trait. Our estimate is that this implies that historically between 10 and 11 percent of children died from malaria or sickle cell disease before reaching adulthood. Such a death rate is roughly twice the current burden of malaria in such regions. Comparing the most affected to least affected areas, malaria may have been responsible for a ten percentage point difference in the probability of surviving to adulthood. In areas of high malaria transmission, our estimate is that life expectancy at birth was reduced by approximately five years. In terms of its burden relative to other causes of mortality, malaria appears to have been perhaps slightly less important historically than it is today, although we certainly can’t rule out the decline in malaria mortality has been proportional to the decline in mortality from other diseases.

Thus, malaria imposed a heavy mortality burden. Did it hold back economic development? We find little reason to believe that it did. Examining the economic burden of malaria mortality in a simple life cycle model suggests that the disease was not very important, primarily because the vast majority of deaths that it caused were among the very young, in whom society had invested few resources. This model-based finding corroborates the findings of our statistical examination. Within Africa, areas with higher malaria burden, as evidenced by the prevalence of the sickle-cell trait, do not show lower levels of economic development or population density in the colonial era data that we examine.

References

- [1] Acemoglu, D., S. Johnson, and J. Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review*, 91, 1369-1401.
- [2] Acemoglu, D. and S. Johnson. 2007. “Disease and Development: The Effect of Life Expectancy on Economic Growth,” *Journal of Political Economy* 115, pp. 925-985,
- [3] Akyeampong, E.l Kwaku. 2006. “Disease in West African History” in Emmanuel Kwaku Akyeampong, ed., *Themes in West Africa’s History*, Athens, OH: Ohio University Press.

- [4] Allison, A. C. 2002. "The Discovery of Resistance to Malaria of Sickle-Cell Heterozygotes." *Biochem. Mol. Biol. Educ.*, 20: 279-87.
- [5] Alsan, M. 2014. "The Effect of the TseTse Fly on African Development," *American Economic Review*. Forthcoming.
- [6] Ashraf, Q. and O. Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*, 101(5), pp. 2003-2041.
- [7] Ashraf, Q., A. Lester, and D. N. Weil. 2008. "When Does Improving Health Raise GDP?" *NBER Macroeconomics Annual* 2008.
- [8] Bairoch, P. 1988. "Cities and Economic Development : From the Dawn of History to the Present." Chicago: University of Chicago Press, 1988.
- [9] Bleakley, H. 2010. "Malaria eradication in the Americas: A retrospective analysis of childhood exposure." *American Economic Journal: Applied Economics*: 1-45.
- [10] Chanda, A. and L. Putterman. 2007. "Early Starts, Reversals and Catch-up in the Process of Economic Development," *Scandinavian Journal of Economics* 109, 387-413.
- [11] Chiovelli, Giorgio, Matteo Cervellati, and Elena Esposito, "Bite and Divide: Malaria and Ethnic Diversity," *Mimeo*, 2015
- [12] Comin, D., W. Easterly, and E. Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?." *American economic journal. Macroeconomics* 2.3: 65-97.
- [13] Cutler, D, W. Fung, M. Kremer, M. Singhal, and T. Vogl. 2010. "Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India." *American Economic Journal: Applied Economics*: 72-94.
- [14] Currat M, Trabucher G, Rees D, Perrin P, Harding RM, et al. 2002. "Molecular analysis of the β -globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the β Senegal mutation". *Am J Hum Genet.* 70:207'223
- [15] Diamond, J. 1997 "Guns, Germs, and Steel: The Fates of Human Societies," New York: W. W. Norton and Co.
- [16] Fenske, J. 2013. "Does Land Abundance Explain African Institutions?" *Forthcoming, Economic Journal*.
- [17] Fenske, J. 2013. "Ecology, trade, and states in pre-colonial Africa," *Forthcoming, Journal of the European Economic Association*.
- [18] Gallup, J. L. and J. D. Sachs. 2001. "The Economic Burden of Malaria," *American Journal of Tropical Medicine and Hygiene*, 64 (1, supplement), pp. 85-96.

- [19] Gallup, J. L., J. D. Sachs, and A. Mellinger. 1999. "Geography and Economic Development," *International Regional Science Review* 22 (2), 179-232.
- [20] Galor, O., and D. N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review* 90, 806-828.
- [21] Gennaioli, N. and I. Rainer, 2007. "The Modern Impact of Precolonial Centralization in Africa" *Journal of Economic Growth* 12, 185-234.
- [22] Greenwood, B., K. Marsh, and R. Snow. 1991. "Why do some African children develop severe malaria?" *Parasitology Today*; 7(10):277-81.
- [23] Hansen, G. D., and E. C. Prescott. 2002. "Malthus to Solow." *American Economic Review*, 92(4): 1205-1217.
- [24] Hedrick, P. W. 2011. "Population Genetics of Malaria Resistance in Humans," *Heredity*, 107 283-304.
- [25] Hill, Adrian V.S. et al. 1991. "Common West Africa HLA antigens are associated with protection from severe malaria," *Nature* volume 352, pp. 595-600.
- [26] Hibbs, D. A., and O. Olsson. 2004. "Geography, biogeography, and why some countries are rich and others are poor," *Proceedings of the National Academy of Sciences* 101, 3715-3720.
- [27] Hagberg, U., and S. Wall. 1986. "Secular Trends in Maternal Mortality in Sweden from 1750 to 1980," *Bulletin of the World Health Organization* 64(1), pp. 79-84.
- [28] Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, J. Sachs, and S. E. Sachs. 2004. "A Global Index of the Stability of Malaria Transmission," *American Journal of Tropical Medicine and Hygiene*, Vol. 70, No. 5, pp. 486-498.
- [29] Kremer, M. 1993. "Population Growth and Technological Change: One Million B.C. to 1990," *Quarterly Journal of Economics* 106:3, 681-716.
- [30] Kurnit D. M. 1979. "Evolution of sickle variant gene." *Lancet*. 1:104.
- [31] Kwiatkowski, D. P. 2005. "How Malaria has Affected the Human Genome and what Human Genetics can Teach us About Malaria," *American Journal of Human Genetics*, 77.
- [32] Lee, R. D. and A. Mason. 2010. "Generational Economics in a Changing World," mimeo.
- [33] Lehmann, H., and A. B. Raper. 1956. "Maintenance of High Sickling in an African Community," *British Medical Journal*, 333-336.

- [34] Livingstone, Frank B. 1958. "Anthropological Implications of Sickle Cell Gene Distribution in West Africa." *Am Anth.* 60: 533-562.
- [35] Lucas, A. M. 2010. "Malaria eradication and educational attainment: evidence from Paraguay and Sri Lanka." *American Economic Journal: Applied Economics* 2.2: 46-71.
- [36] Luzzatto L. Sickle Cell Anaemia and Malaria. *Mediterranean Journal of Hematology and Infectious Diseases.* 2012;4(1):e2012065. doi:10.4084/MJHID.2012.065.
- [37] Mabogunje, A. L. and P.I Richards. 1985. "The land and peoples of West Africa" in Ajayi, J.F.A. and Michael Crowder, eds. *History of West Africa*, volume 1, 3rd edition.
- [38] Maleney, P. , A. Spielman, and J. D. Sachs. 2004, "The Malaria Gap," *Am J. trop hygiene and medicine*;71(2 Suppl):141-6.
- [39] McCann, J. (2011) "Deposing the Malevolent Spirit: A Historical Cultural Ecology of Malaria in Northwest Ethiopia". *PSAE research series Issue 9*. African Studies Center, Boston University.
- [40] McGuire, R. A., and P. R. P. Coelho. 2011. "Parasites, Pathogens, and Progress: Diseases and Economic Development," Cambridge, MA: MIT Press.
- [41] McNeill, W. 1977. "Plagues and Peoples, New York: Anchor Books.
- [42] Morrow, R. H., and W. J. Moss. 2006. "The Epidemiology and Control of Malaria" in Nelson, Kennard, and Carolyn M. Williams, eds., *Infectious Disease Epidemiology: Theory and Practice* second edition, Jones and Bartlett Publishers, 1087-1138.
- [43] Motulsky, A. 1964. "Hereditary Red Cell Traits and Malaria," *Am. J. Trop. Med. Hyg.*, 13 (Part 1), pp. 147-158
- [44] Murdock, G. R. "Africa: Its people and their history." McGraw-Hil, NY (1959).
- [45] Murdock, George Peter. "Ethnographic atlas." (1967).
- [46] Murray, C. J. D., and A. D. Lopez. 1996. "The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020, " Cambridge, MA: Harvard University Press.
- [47] National Center for Atmospheric Research Staff (Eds). 2014. "The Climate Data Guide: CRU TS3.21 Gridded precipitation and other meteorological variables since 1901"

- [48] National Heart, Lung, and Blood Institute. 1996. "Facts About Sickel Cell Anemia." NIH Publication No. 96-4057.
- [49] Nelson, Kennard and Carolyn M. Williams. 2006. *Infectious Disease Epidemiology: Theory and Practice*. 2nd Edition. Jones and Bartlett Publishers.
- [50] Persad, G., A. Wertheimer, and E.I J. Emanuel. 2009. "Principles for allocation of scarce medical interventions," *The Lancet*; 373: 423-31.
- [51] Piel, F.B. et al. 2010. "Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis." *Nat. Commun.* 1:104 doi: 10.1038/ncomms1104.
- [52] Putterman, L., and D. N. Weil. 2010. "Post-1500 population flows and the long-run determinants of economic growth and inequality." *The Quarterly Journal of Economics* 125.4: 1627-1682.
- [53] Ramankutty, N., J. A. Foley, and J. Norman and K. McSweeney. 2002. "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change," *Global Ecology and Biogeography*, 11, 377-392.
- [54] Riley, J. C. 2005. "Estimates of Regional and Global Life Expectancy, 1800-2001," *Population and Development Review* 31(3), 537-543.
- [55] Soloman E., and W.F. Bodmer.1979. "Evolution of sickle variant gene." *Lancet*. 1:923.
- [56] Steyn, M. 2003."A Comparison between Pre- and Post-Colonial Health in the Northern Parts of South Africa, a Preliminary Study," *World Archaeology*, Vol. 35:2, *Archaeology of Epidemic and Infectious Disease*, pp. 276-288
- [57] Stokey, Nancy L. 2001. "A Quantitative Model of the British Industrial Revolution, 1780-1850." *Carnegie-Rochester Conference Series on Public Policy*, 55: 55-109.
- [58] Stone, L. 1977. "The Family, Sex and Marriage in England, 1500-1800,".
- [59] Spolaore, E., and R. Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, *American Economic Association*, vol. 51(2), pages 325-69, June.
- [60] United Nations, Department of Economic and Social Affairs, Population Division. 2009."World Population Prospects: The 2008 Revision", New York.
- [61] United Nations, Department of Economic and Social Affairs, Population Division.1999. "The World at Six Billion" New York.
- [62] United Nations, Department of Economic and Social Affairs, Population Division. 1982. "Model Life Tables for Developing Countries."

- [63] United Nations Department of Economic and Social Affairs, Population Division. 1983 “Manual X: Indirect Techniques for Demographic Estimation” (United Nations publication, Sales No. E.83.XIII.2).
- [64] Voigtlander, N., and H-J. Voth. 2013. “The three horsemen of riches: Plague, war, and urbanization in early modern Europe.” *The Review of Economic Studies* 80.2: 774-811.
- [65] Webb, J. L. A. 2006. “Ecology and Culture in West Africa” in Emmanuel Kwaku Akyeampong, ed., *Themes in West Africa’s History*, Athens, OH: Ohio University Press.
- [66] Weil, D. N. 2010. “Endemic Diseases and African Economic Growth: Challenges and Policy Responses,” *Journal of African Economies*.
- [67] Weil, D. N. “The Impact of Malaria on African Development over the Longue Duree” in Emmanuel Akyeampong, Robert Bates, Nathan Nunn, and James A. Robinson eds., *Africa’s Development in Historical Perspective*. Cambridge University Press, 2014.
- [68] Weil, D. N. and J. Wilde. 2009. “How Relevant is Malthus for Economic Development Today?” *American Economic Review* 99(2): 255-60.
- [69] Weisenfeld, S. L. 1967. “Sickle Cell Trait in Human Biological and Cultural Evolution,” *Science*, CLVII, 1135-1140.
- [70] Williams T.N., T.W. Mwangi, S. Wambua, N.D. Alexander, M. Kortok, R.W. Snow, and K. Marsh. 2005. “Sickle cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases,” *Journal of Infectious Diseases*, 192(1):178-86.
- [71] World Health Organization. 2014. *World Malaria Report*.

FIGURES AND TABLES

Table 1: Components of the Cost of Malaria

Group	Fraction of Births	Death Rate from Malaria or Sickle Cell Disease	Fraction of all Children who Die in Category
Non-Carriers (AA)	$(1 - \frac{\pi}{2})^2$	M^{AA}	$(1 - \frac{\pi}{2})^2 M^{AA}$
Carriers (AS)	$\pi (1 - \frac{\pi}{2})$	M^{AS}	$\pi (1 - \frac{\pi}{2}) M^{AS}$
Sickle Cell Disease (SS)	$(\frac{\pi}{2})^2$	1	$(\frac{\pi}{2})^2$

Table 2: OLS Estimates: Historical Malaria Burden and Population Density
Dependent Variable: Log of Population Density (Alsan 2014)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Malaria Burden	0.236*** (0.053)	0.247*** (0.049)	0.228*** (0.051)	0.182*** (0.069)	0.179*** (0.069)	0.185*** (0.07)	
Highly Malarious Environment c.1900							0.329*** (0.082)
Geographic Controls	Y	Y	Y	Y	Y	Y	Y
Waterways Controls	N	Y	Y	Y	Y	Y	Y
Climate Controls	N	N	Y	Y	Y	Y	Y
Cultural Cluster Dummies	N	N	N	Y	Y	Y	Y
Suitability Slash and Burn Agriculture	N	N	N	N	Y	Y	Y
Tse Tse Suitability	N	N	N	N	N	Y	Y
Observations	398	398	398	398	398	398	398
R-squared	0.168	0.234	0.238	0.499	0.499	0.500	0.522

Standardized beta coefficients. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Geographic controls are mean value of the index of land suitability for agriculture (Ramankutty et al., 2002), ecological diversity (Fenske, 2013), and mean elevation of terrain. Waterways controls are the log total area of water bodies accessible for the ethnic and the log of the geodesic distance of the centroid of the historical homeland of each ethnic group from the nearest coastline. Climate controls are mean and standard deviation of temperature calculated for each ethnic group homeland over the time period 1901 - 2012 (using CRU TS3.21 dataset). Cultural cluster dummies account for 41 cultural provinces in the Ethnographic Atlas. Suitability for Slash and Burn Agriculture is the mean value of the suitability index for cultivation of yams and bananas (FAO's Global Agro-Ecological Zones project -GAEZ- v3.0). Tse-tse Suitability is an index from Alsan (2014).

Table 3: OLS Estimates: Historical Malaria Burden and Other Measures of Prosperity

	Dependent Variable			
	> 20k City in 1850	Settlement Complexity	Centralization of Power	Intensive Agriculture
	(1)	(2)	(3)	(4)
Malaria Burden	0.049 (0.054)	0.147* (0.077)	0.102 (0.073)	-0.079 (0.067)
Observations	520	482	520	520
R-squared	0.190	0.359	0.216	0.352

Standardized beta coefficients. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.
All specifications include the full set of controls from Table 2.

Table 4: First-Stage: Plasmodium Suitability versus Malaria Ecology

	Dependent Variable: Historical Malaria Burden					
	(1)	(2)	(3)	(4)	(5)	(6)
Plasmodium Suitability	0.48*** (0.075)	0.41*** (0.079)			0.46*** (0.076)	0.40*** (0.078)
Malaria Ecology			0.18*** (0.064)	0.07 (0.053)	0.08 (0.053)	0.04 (0.059)
Climate Controls	N	Y	N	Y	N	Y
Other Controls in Table 2	Y	Y	Y	Y	Y	Y
Observations	520	520	520	520	520	520
R-squared	0.744	0.753	0.716	0.735	0.745	0.754

Standardized beta coefficients. Robust standard errors in parentheses *** p<0.01, ** p<0.05,
, * p<0.1. All specifications include the full set of controls from Table 2.

Table 5: IV Estimates: Historical Malaria Burden and Pre-Colonial Development

	Dependent Variable				
	Log of Population Density	> 20k City in 1850	Settlement Complexity	Centralization of Power	Intensive Agriculture
	(1)	(2)	(3)	(4)	(5)
Malaria Burden	-0.059 (0.354)	-0.597 (0.377)	0.208 (0.259)	-0.180 (0.283)	-0.368 (0.278)
F-Statistic	21.809	26.893	24.372	26.893	26.893
Observations	398	520	482	520	520

Standardized beta coefficients. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.
All specifications include the full set of controls from Table 2.

Table 6: Further Robustness Checks (IV Estimates)

Dependent variable	(1)	(2)	(3)	(4)	(5)
Log of Population Density	-0.083 (0.362) [18.9]	-0.067 (0.479) [11.04]	-0.006 (0.369) [20.23]	0.001 (0.515) [9.1]	-0.361 (0.472) [13.72]
> 20k City in 1850	-0.519 (0.399) [23]	-0.960 (0.595) [14.9]	-0.558 (0.394) [25.5]	-0.756 (0.608) [12.83]	-0.859* (0.468) [17.83]
Settlement Complexity	0.134 (0.282) [20.5]	0.289 (0.375) [15.6]	0.193 (0.269) [22.64]	0.157 (0.403) [13.46]	0.263 (0.298) [17.13]
Centralization of Power	-0.130 (0.307) [23.04]	-0.155 (0.433) [14.9]	-0.144 (0.289) [25.5]	-0.029 (0.464) [12.83]	-0.169 (0.325) [17.83]
Intensive Agriculture	-0.439 (0.302) [23.04]	-0.508 (5.038) [14.9]	-0.359 (0.287) [25.5]	-0.584 (0.471) [12.83]	-0.307 (0.319) [17.83]
Robustness	External Influence Controls	Shortest Distance to Origin of Mutation	HbC Mutation	All Previous Controls	Only Tropical Sample

Standardized beta coefficients. Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. First Stage F-Statistics in Brackets. External influence controls are the log of (1+ total slaves exports normalized by area of homeland) from Nunn and Wantchekon (2011), and two dummy variables indicating if European explorers traveled (between 1768 and 1894), or a historical trade route passed through, the homeland of the ethnic group. Shortest distance to origin of mutation captures the shortest distance from each ethnic homeland to the closest hypothesized origin of the mutation. HbC Mutation is the average prevalence of the Hemoglobin C mutation (from MalariaAtlas Project). Tropical sample includes only ethnic group with homelands within the tropics ($N = 483$).

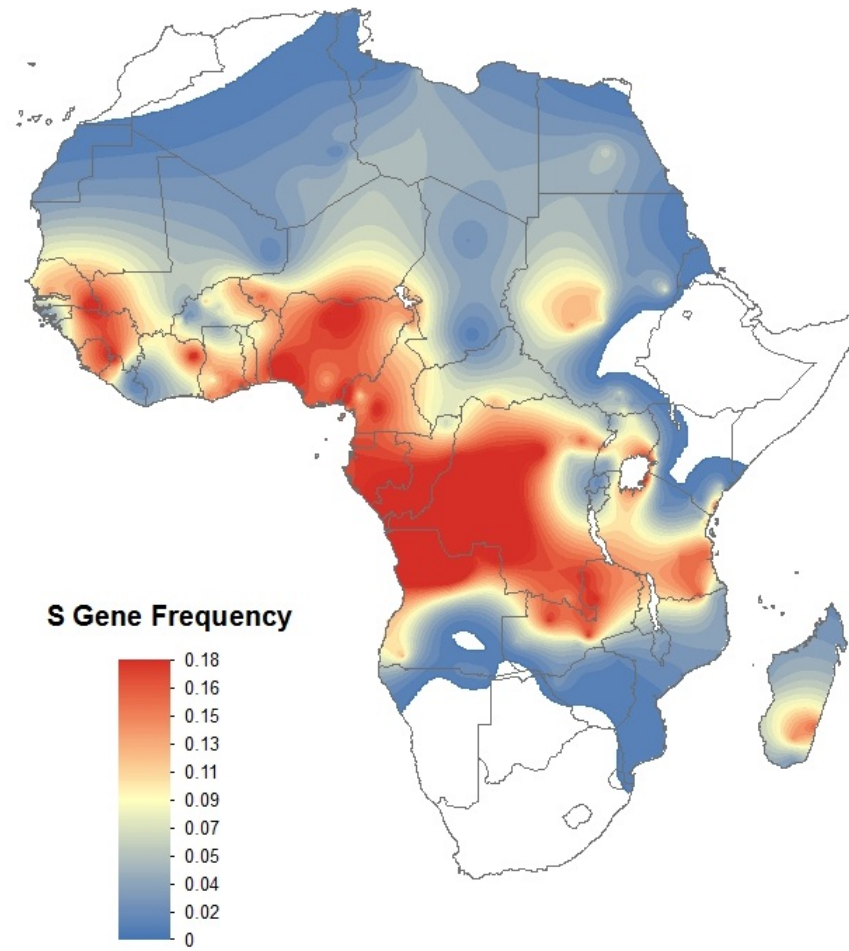


Figure 1: Sickle Cell Gene Frequency from Piel et al (2010)

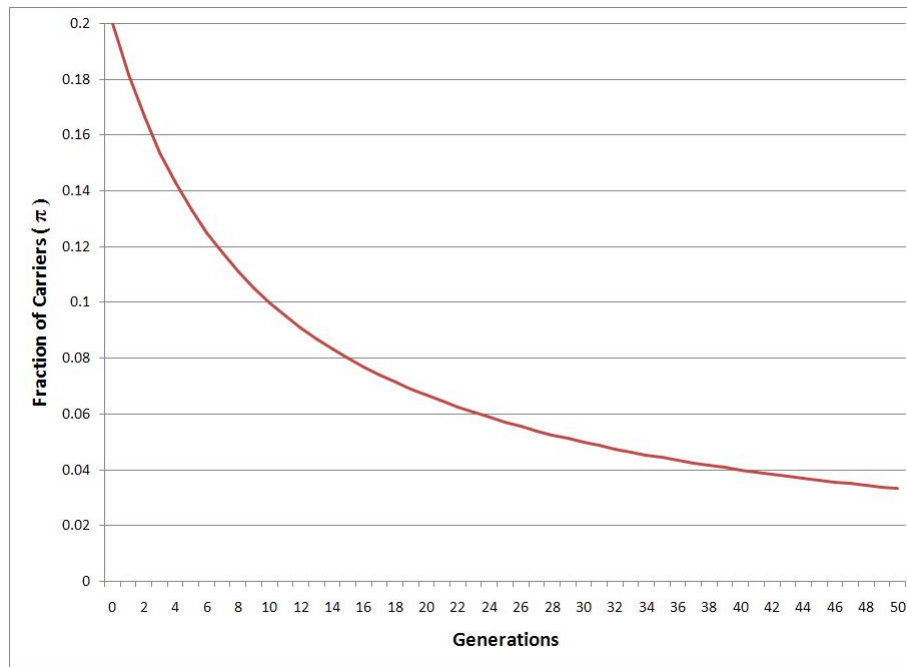


Figure 2: Effect of a Hypothetical Malaria Eradication on Evolution of Carriers

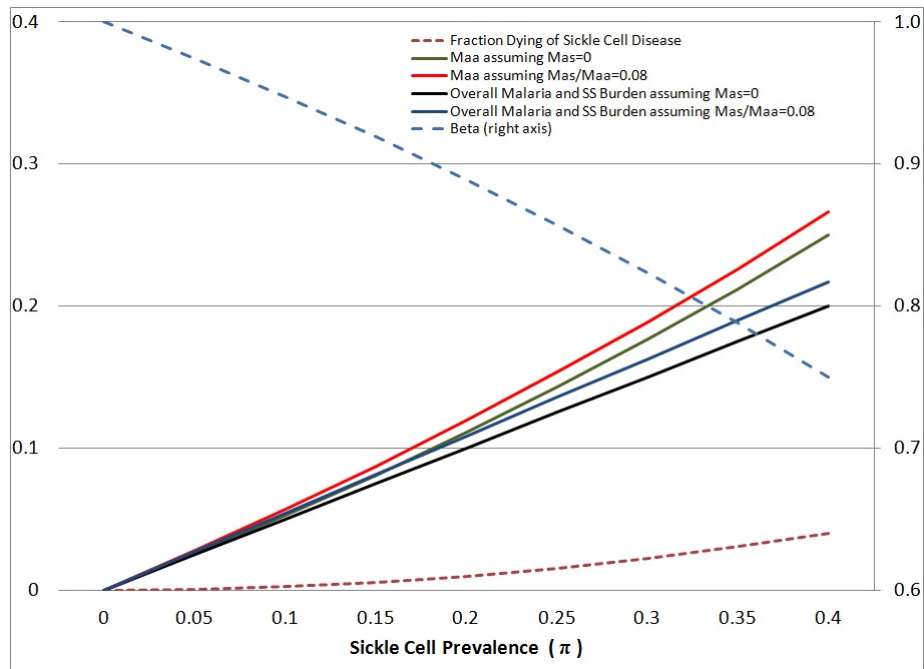


Figure 3: Implications of Varying Sickle Cell Prevalence for Malaria Burden

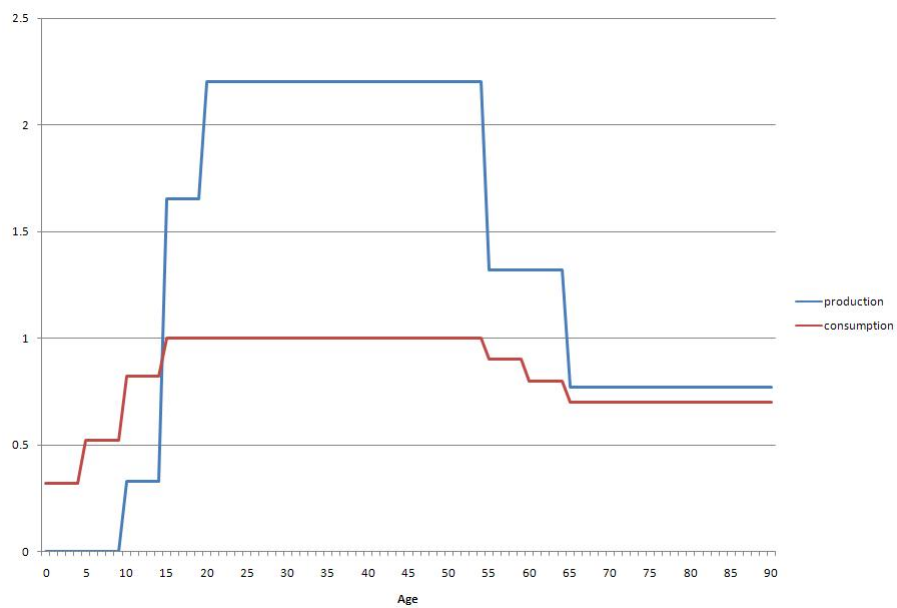


Figure 4: Consumption and Income Profiles

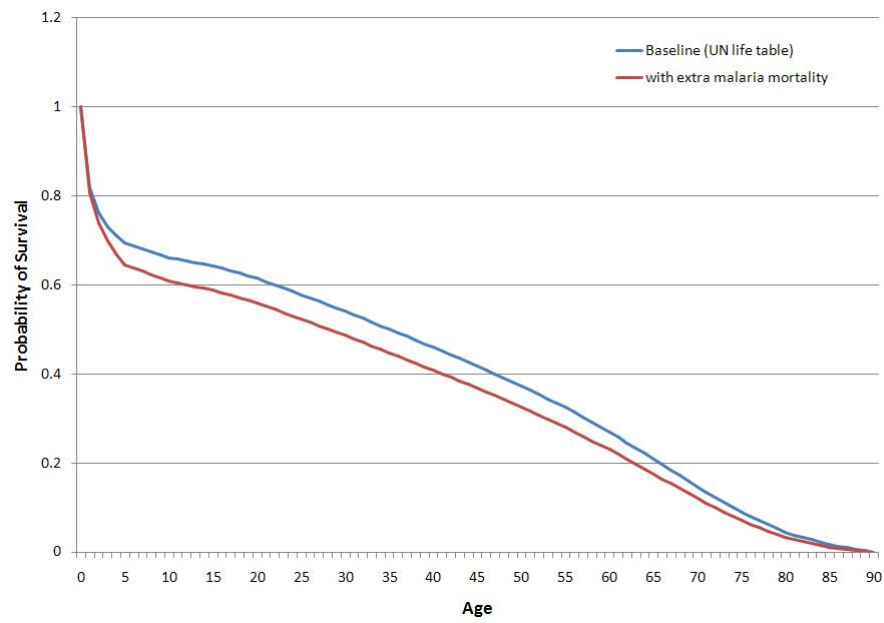


Figure 5: Change in Survival Probabilities due to Malaria