WHAT CAN THE HUMAN GENOME TELL US ABOUT THE ECONOMIC EFFECTS OF MALARIA?

Patricia Jones
University of Oxford
This Version: June 2018

ABSTRACT

This paper proposes a new instrumental variable for malaria incidence which exploits the regional variation in genetic mutations. The source of identification is the heterogeneity in rates of G6PD—glucose-6-phosphate dehydrogenase—deficiency across countries. Rates of G6PD deficiency are highest in countries with a history of malaria transmission. Most importantly for the purposes of this study, the vast majority of individuals who inherit G6PD deficiency are asymptomatic throughout their life. By exploiting differences in the prevalence of G6PD deficiency across countries, I find strong evidence that malaria is causally linked to lower per capita income.

Keywords: Health, Economic Development, Economic Growth

JEL Classification: I15, I130, O1, O4

I. Introduction

Malaria is the world's most prevalent vector-borne disease. Currently, there are 99 countries with on-going malaria transmission and another five where malaria is likely to make a comeback if further preventative efforts are not taken (See Figure 1). Approximately half of the world's population live in a malarias zone where more than 200 million clinical cases of malaria are reported each year (WHO, 2012). Globally, malaria claims about 600,000 lives per year, the vast majority of these deaths occur in children who are less than five years old (WHO, 2012). In Sub-Saharan Africa, transmission rates are so high that it is estimated that one child dies every minute from the disease (WHO, 2013).

Since the early 2000s there has been a large global effort to reduce malaria transmission. International disbursements for malaria control rose from less than 100 million US dollars in 2000 to more than 1.5 billion US dollars in 2010 (WHO, 2012, p. 15). While the primary justification for such funding was to reduce the amount of suffering and premature death caused by malaria, there was a secondary goal as well—to stimulate the rate of economic growth in malaria endemic countries. Theoretically, health investments should produce large economic benefits by increasing worker productivity. Empirical evidence from microeconomic studies largely supports this causal link (see, for example, Grossman, 2000; Strauss and Thomas, 1997; Case and Paxson, 2002; and Schultz, 2002). By contrast, recent studies which have measured the impact of health improvements on macroeconomic outcomes have found little or no significant effects (Acemoglu and Johnson, 2007; Weil, 2007; Ahraf, Lester, and Weil, 2009).[1]

Empirical evidence at the macro level is weak because of the difficulties in measuring the direct effects of health on aggregate output. The main problem is that health (or disease incidence) is likely to be an endogenous variable. Countries with poor health tend to be disadvantaged in other ways as well which affect both their ability to control the spread of disease and their capacity to generate economic growth. Cross-country regressions which include health outcomes as independent variables may be capturing the effects of these other (often omitted) disadvantages rather than the effects of health itself. Solving the endogeneity problem is

[1] Several older studies have found a positive and significant effect of increased life expectancy on economic growth. See Bloom et al, 2004 for a survey of this literature.

not an easy task because it is difficult to find an exogenous source of variation in health which is not correlated with the error term in a growth equation.

This study proposes a new instrumental variable for malaria which exploits the regional variation in genetic mutations which have arisen through natural selection in order to protect individuals against malaria. The source of identification is the heterogeneity in rates of G6PD—glucose-6-phosphate dehydrogenase—deficiency across countries. G6PD deficiency is the most common human enzyme deficiency in the world (see Figure 2), affecting an estimated 400 million people worldwide (Weatherall, 2008). Rates of G6PD vary widely across countries but are highest in regions with a history of malaria transmission. Most importantly for the purposes of this study, the vast majority of individuals who inherit G6PD deficiency remain asymptomatic throughout their life. Medical researchers claim that "G6PD deficiency does not seem to affect life expectancy, quality of life, or the activity of affected individuals" (Capellini and Fiorelli, 2008, p. 68). Moreover, the protective effects of G6PD deficiency diminish as a person ages so that, by the time an individual reaches adulthood, he (or she) has developed his (or her) own immunity against malaria from prior infections.

Clinical studies suggest that only a small proportion of the overall variation in malaria resistance in *adults* can be attributed to the most well-known genetic mutations (Hedrick, 2011). Genetics plays a larger role in protecting young children from malaria because they have not yet acquired immunity from exposure to the disease. How much G6PD deficiency confers resistance is still a matter of debate. Clinical studies show that G6PD deficiency is linked to increased resistance in males (but not females) to severe forms of malaria (Johnson et al, 2009). The evidence is mixed for mild cases of malaria which are more common. The most recent research finds that G6PD deficiency does not increase resistance to uncomplicated cases of malaria (Johnson et al, 2009) but other studies have found increased resistance in females (Ruwende et al, 1995; Guido et al, 2007). The level of increased resistance in children varies from 0% to about 50% but declines with age (Kendrix, 2011). Therefore, I assume that any correlation between G6PD deficiency and increased malaria resistance among the *adult* population is small and will not affect the validity of my estimated coefficients.

In this study, I adopt the following identification strategy: G6PD deficiency is correlated with contemporary measures of malaria but not correlated to other country characteristics which

could potentially affect economic performance. There are several possible weaknesses with this identification strategy. First, the genetic disorder is more prevalent among males than it is among females because of its pattern of X-linked inheritance.[2] To correct for this imbalance, the estimates of G6PD deficiency are adjusted for differences in the sex-ratios across countries. Second, G6PD deficiency is present in populations which have inherited other "loss of function" genetic disorders (e.g., sickle cell anaemia). Consequently, it is possible that the instrumental variable is picking up the effects of these other omitted diseases. To address this possibility, I estimate separate regressions which include country-level estimates of sickle-cell anaemia as a covariate. Third, there is some evidence that G6PD deficiency is correlated with increased protection against severe malaria in African children. If such protection is also correlated with greater human capital accumulation, the instrumental variable may be capturing this effect. Therefore, I include as a covariate a measure of human capital. Lastly, the results are based on cross-country regressions which limit my ability to control for unobserved differences across countries that are potentially correlated with malaria incidence. This weakness remains given my identification strategy.

Figure 3 shows a strong, positive correlation between rates of G6PD deficiency at the country-level and levels of malaria incidence. It plots the percentage of a country's population with G6PD deficiency against the percentage of a country's population at stable risk of contracting *Plasmodium falciparum*—the most deadly form of the malaria parasite—for all malaria endemic countries. The (first stage) regression reveals that rates of G6PD deficiency explain about 40% of the variation across countries in the population at risk (PAR) of *P. falciparum*. Figure 4 reveals a similar relationship. It plots the prevalence of G6PD deficiency against the logarithm of disability-adjusted life-years (DALYs) which are lost as a result of malaria.[3] The graph reveals a strong, positive correlation between average DALYs lost due to malaria and the incidence of G6PD deficiency.

[2] The Malaria Atlas Project explains the genetics of G6PD deficiency as follows: "The gene is found on the X chromosome. Women have two copies of the X chromosome and therefore two opportunities to inherit a copy of the gene which codes for normal levels of G6PD. Men only have one copy of the X chromosome and therefore only one copy of the G6PD gene. Men are therefore more likely to suffer from severe G6PD deficiency" (See http://www.map.ox.ac.uk/explore/inherited-blood-disorders/g6pd-deficiency).

[3] DALYs estimate the years of healthy life lost by a person due to disease, disability, or death. For this study, I use the average DALYs lost per person as a result of malaria.

The results from my analysis reveal that malaria has a significant impact on aggregate output. Specifically, for every 1% increase in the proportion of population who are at risk of contracting *P. falciparum*, average output per worker falls by 0.5%, after controlling for the standard production and geography variables. While this effect may seem small, the economic benefits associated with malaria eradication are substantial. For example, if *P. falciparum* were totally eradicated (that is, the population at stable risk of *P. falciparum* were reduced to zero), average output per worker would rise by about 50%. These results complement those found by Gallop and Sachs (2001a, 2001b) and others (Cartensen and Grundach, 2006; Bhattacharya, 2009) who argue that malaria has had a negative impact on long-run economic development. My results show that malaria continues to lower per capita income in malaria endemic countries, even after controlling for a country's current level of capital per worker, human capital, and geographic characteristics.

The study contributes to the current economic literature in two primary ways. First, it proposes a new instrumental variable for malaria incidence which is not based on either geographic characteristics or the timing of malaria eradication campaigns. Indeed, it is the first study to my knowledge that uses natural selection as an exogenous source of variation for disease prevalence. Second, my study contributes to the broader debate on the role of health in determining income differences across countries (Bloom et al, 2004; Acemoglu and Johnson, 2007; Weil, 2007; Ashraf, Lester, and Weil, 2009). There are many possible channels through which improved health status may affect a country's output. Human capital theory predicts that health investments have large economic payoffs in terms of higher worker productivity and earnings (Grossman, 2000). Numerous studies reveal that healthier children grow-up to be taller and more productive adults (Steckel, 1995; Case, Lubotsky, and Paxson, 2002; Case and Paxson, 2008). Healthier children miss fewer days from school and learn more during the days which they attend school (Miguel and Kremer, 2004; Alderman, Hoddinott, and Kinsey, 2006; Bleakley, 2007; Maluccio et al, 2009). Whether or not such microeconomic effects produce significant effects at the aggregate level, however, remains a topic of debate.[4]

[4] See Packard (2009) for a good summary of the difficulties in measuring the macroeconomic effects of malaria control from microeconomic studies.

The remainder of this paper is organized as follows. Section 1 provides a survey of the economic literature which examines the link between malaria eradication and economic outcomes. Section 2 discusses the current literature on malaria genetics and how it can be used to test the economic effects of malaria. Section 3 describes the model tested and the data sources used to test the model. Section 4 presents the ordinary least squares (OLS) estimates between malaria incidence and economic development. In addition, it presents the main results from the two-stage least squares (2SLS) regression analysis and several robustness checks. Finally, Section 5 includes some final remarks and suggestions for future research.

2. Background on the Economic Effects of Malaria

Since the early 2000s, global funding targeted at malaria prevention has increased by a factor of about 15. One rationale for this funding was the belief that improved health status would have a large, positive impact on economic growth. The best known advocate of this view is Jeffrey Sachs who wrote a series of papers with several co-authors which argued that malaria is largely responsible for the poor economic performance of Sub-Saharan Africa (Gallup and Sachs, 2001a, 2001b; Sachs and Maleney, 2002; Sachs, 2003). In a much quoted finding, Sachs and Malaney (2002) state: "countries in which a high proportion of the population lived in regions of *P. falciparum* malaria transmission in 1965 had annual growth rates that were 1.3% lower than other countries over the period 1965-1990" (p. 681).

Since Sachs's influential research, several studies have re-examined the potential impact of malaria on economic outcomes. Much of this research relies on historical data and cohort studies. The identification strategy employed in these studies has been to use geographical variations in pre-eradication malaria risk both within a country and across time to measure the difference in economic outcomes among cohorts with greater exposure to malaria. One advantage of this type of research is that historical records can be used to measure how malaria exposure during childhood (or even while in utero) affects adult outcomes.

Barreca (2010), for example, matches adults in the 1960 US census to the state-level malaria death rate in their respective year and state of birth and then instruments for malaria exposure using the within-state variation in "malaria ideal temperatures." He finds that greater exposure to malaria risk while in utero and during postnatal periods significantly reduces

educational attainment and "can account for as much as 25 percent of the difference in long-term educational attainment between cohorts born in malaria-afflicted states and cohorts born in nonafflicted states during the early 20th century" (p. 868). Similarly, Chang et al. (2011) estimate the long-run impact of malaria exposure around birth in colonial Taiwan and find that early exposure to malaria leads to lower levels of educational attainment and worsened adult health outcomes.

Several other studies have focused on the link between malaria eradication and human capital accumulation. For example, Bleakley (2010) estimates the long-run impact of childhood exposure to malaria in the United States, Brazil, Colombia and Mexico. He finds strong evidence that cohorts born after the anti-malaria campaigns had higher adult literacy than those born before the campaigns. Lucas (2010) conducts a similar analysis for Paraguay and Sri Lanka. She finds evidence that malaria eradication led to increased female education attainment for both countries. The empirical evidence, however, does not always support a positive link between malaria eradication and higher levels of educational attainment. Cutler et al (2010), who examine the impact of malaria eradication on education in India, "find no evidence of increased educational attainment for men and mixed evidence for women" (p. 72). Similarly, Venkatarami, (2012) finds no evidence that the anti-malaria program carried out in Mexico increased educational attainment.

These last two studies should not be interpreted as evidence that malaria eradication failed to increase human capital, even though it failed to increase average levels of educational attainment. It is possible to increase human capital without achieving higher levels of education if the quality of education improves or if students learn more during each year of education. This appears to be what happened in Mexico. In Mexico, there is evidence that cohorts who were born after the anti-malaria campaigns started school earlier and scored higher on cognitive tests. In addition, both Cutler and Venkatarami find evidence that, at least for men, household consumption increased for those cohorts born after malaria eradiation. Therefore, the micro evidence generally supports positive links between malaria eradication, human capital, and worker productivity.

3. Malaria and Human Genetics

Today it is widely understood that specific gene variants can contribute to disease resistance in humans (Carter, 2002; Dronamraju, 2004; Weatherall, 2008). Numerous genetic mutations, for example, have been linked to malaria and are believed to increase resistance against the disease. As described by one scientist: "Malaria is the evolutionary driving force behind sickle-cell disease, thalassemia, glucose-6-phosphatase deficiency, and other erythrocyte defects that together comprise the most common Mendelian diseases of humankind" (Kwiatowski, 2005). While scientists still do not fully understand how this resistance mechanism works and whether all gene mutations confer resistance, their understanding of the genomic epidemiology of malaria has increased significantly over the past decade.

Scientific research has revealed, for example, that different populations have developed independent evolutionary responses to malaria. The HbS allele (responsible for sickle-cell anaemia) is common in Africa but not in Southeast Asia while the HbE allele (responsible for β-thallasemia) is common in Southeast Asia but not Africa (Kwiatowski, 2005). Even within Africa, there is evidence that the HbS allele evolved independently in four different populations. Recently, about 140 mutations in the G6PD gene have been identified although most of these (121 out of 140) are amino acid substitutions rather than severe mutations (Howes et al, 2013).

G6PD deficiency is found mainly in Africa, Asia, and Mediterranean Europe. Scientists date the natural selection process which produced G6PD deficiency somewhere between 3,840 and 11,760 years ago (Tishkoff et al, 2001).[5] Those who inherit G6PD deficiency are more at risk of developing haemolysis (the destruction of red blood cells) because their red blood cells have a diminished ability to withstand oxidative stress. Haemolysis occurs as a result of exposure to external factors like certain drugs or infections (Cappellini and Fiorelli, 2008). The risk of developing haemolysis, however, varies by the severity of G6PD deficiency. There are four broad classes of G6PD deficiency: class I (patients have severe mutations with haemolytic anaemia); class II (patients have < 10% of normal G6PD function); class III (patients have 10-60% of normal G6PD function); and class IV (patients have 60-100% of normal G6PD function). Class I deficiencies are extremely rare and can cause transfusion dependency whereas individuals who inherit Class II, III, and IV deficiency are usually asymptomatic. Class III is the most common form of G6PD deficiency (Peters and Van Noorden, 2009).

[5] There is no evidence that G6PD deficiency confers resistance to other topical diseases such as yellow fever, sleeping sickness, or dengue fever.

As stated in the introduction, the vast majority of people who inherit G6PD deficiency remain asymptomatic throughout their life.[6] Because G6PD deficiency poses little health risk, limited data existed on its prevalence until recently. Interest in the disease grew after it was discovered that some individuals with G6PD deficiency have a severe drug reaction to primaquine—the most common drug used to treat *P. vivax* malaria. This discovery led to a demand for more accurate data on the distribution of G6PD deficiency. In 2012, researchers affiliated with the Malaria Atlas Project (MAP)[7] responded by generating a map of the global distribution of G6PD deficiency (see Figure 2). The estimates used to construct this map are based on a Bayesian statistical model which was specifically adapted to the gene's X-chromosome inheritance pattern.[8] Using the information contained in this map, a country-specific estimate of the prevalence of G6PD deficiency can be obtained for each malaria endemic country. There are no equivalent data for malaria-free countries. Scientists report that rates of G6PD deficiency are very low in these countries (Nkoma et al, 2009). Therefore, I assign a value of zero to G6PD deficiency for non-malaria endemic countries. All regressions include a dummy variable which takes the value of one if the country is malaria-free which controls for these missing values.

According to Howes (2012), the highest frequency of G6PD deficiency occurs in currently malarious regions and countries with a history of malaria transmission. National frequency estimates range from 0.1% in Cape Verde and the Democratic People's Republic of Korea to 23% in Benin (Howes et al, 2012). On a regional basis, the highest frequencies are found in Sub-Saharan Africa. However, the largest numbers of individuals living with G6PD deficiency reside in Asia. Approximately, 40% of all males with G6PD deficiency live in China and India while 28% live in Sub-Saharan Africa and 4.5% live in the Americas (Howes et al,

[6] This is not the case for other malaria-linked genetic diseases. The two other classic examples of natural selection in response to malaria are sickle-cell anaemia and thalassemia. Sickle-cell is caused by a variation in the gene which codes for haemoglobin (the protein in our red blood cell that helps to carry oxygen from our lungs to the rest of our body). This gene is known as the HbS gene. People with sickle-cell are born with it and have inherited the HbS gene from both parents. While symptoms vary, people with sickle-cell anaemia usually have chronic pain, fatigue, and a reduced life expectancy. People with sickle cell *trait* are different. They have inherited the gene from only one parent and do not have the disease. Instead, they are believed to have increased resistance against malaria. Like sickle-cell anaemia, thalassemia causes the body to make fewer healthy, red blood cells. People with thalassemia have mild to severe anaemia.

[7] The Malaria Atlas Project (MAP) is a research group comprised of two teams of researchers—one based at the Spatial Ecology and Epidemiology Group (SEEG) at the University of Oxford and the other based at the Centre for Geographic Medicine in Nairobi, Kenya.

[8] See Howes et al (2012) for an explanation of how the numbers of deficient individuals were estimated.

2012). The health risks associated with G6PD deficiency also vary by region because of genetic differences associated with the evolution of the G6PD allele. The lowest risk of severe haemolysis exists in Sub-Saharan Africa (despite its high prevalence of the disease) while those with the highest risk of severe haemolysis are found in the Arabian Peninsula and across West Asia. Overall, it is estimated that about 8% of the total population in malaria-endemic regions have G6PD deficiency (Howes et al, 2012).

In this paper, I use national estimates of G6PD deficiency generated by Howes et al (2012) as an instrumental variable for malaria. My identification strategy is based on the assumption that rates of G6PD deficiency are correlated with contemporary measures of malaria but not correlated to other country characteristics which could potentially affect economic performance. Specifically, I make the following two assumptions: (1) past levels of malaria incidence have produced evolutionary pressures which led to genetic variations in the G6PD gene; and (2) these genetic variations are correlated to contemporary measures of malaria incidence and DALYs lost from malaria.

During the past century, more than 50 countries have eradicated malaria (Tatem et al, 2013). Thus, assumption (2) is unlikely to hold unless I control for the existence of past malaria vectors. Therefore, all regressions include a control for whether the country has a history of malaria and the number of years that the country has been malaria free. In addition, assumption (2) is unlikely to hold if past migration has weakened the link between estimates of G6PD deficiency and contemporary measures of malaria. During the past 10,000 years (the period since the first malaria-linked genetic mutations) there have been large migrations of human populations across different regions. Malaria is often transported from one place to another by migrating individuals who carry the malaria parasites in their blood. Depending upon the strain of malaria, these parasites can live in a person anywhere from two years (*P. falciparum*) to four years (*P. vivax*) and even as long as fifty years for some strains (*P. malaria*).[9] While it is impossible to completely control for the effects of such migration, I include in my regressions a measure of the level of ethnic fractionalization as a rough measure of the heterogeneity of the gene pool within each country.

[9] Montavalvo and Reynal-Querol (2007) find evidence that "13% of the cases of malaria reported by the WHO are caused by forced migration as a consequence of civil war" (p. 165).

4. Data and Empirical Strategy

The hypothesis being tested in this paper is that higher levels of malaria incidence and DALYs lost from malaria are causally linked to lower levels of GDP per capita. To test this hypothesis, I estimate the following equation:

$$lny_c = \alpha + \delta \ln k_c + \gamma lnh_c + \beta Malaria_c + \boldsymbol{X_C^G \Pi} + \varepsilon_c \quad (1)$$

where lny_c is the log of output per worker in 2010 expressed in PPP, c denotes the country, lnk_c is the log of capital-labour in 2010 expressed in PPP, lnh_c is the log of human capital, $Malaria_c$ is one of three measures of malaria incidence or malaria-related morbidity, $\boldsymbol{X_C^G}$ is a vector of ethnic and geographic control variables and ε_c is an error terms with zero mean and common variance.[10] To address the potential problems of endogeneity, I employ a two-stage least squares estimation procedure. In the first-stage, $Malaria_c$ is regressed on all the exogenous variables. That is,

$$Malaria_c = \tau + \rho G6PD_c + \varphi \ln k_c + \omega lnh_c + \boldsymbol{X_C^G \Pi} + \varepsilon_{Malaria} \quad (2)$$

where $G6PD_c$ refers to the percentage of the population in country c who have inherited G6PD deficiency. The exclusion restriction is that $G6PD_c$ does not appear in equation (1).

To estimate equations (1) and (2), I construct a cross-country data set for 119 countries which includes a full set of production data, four malaria measures, and several geographic covariates. All data are from published sources and, when possible, represent the most recent data available. Production data are obtained from the Penn World Tables (PWT), Version 8.[11] In this version of the PWT, a new measure of GDP has been introduced which is constructed to better reflect the production possibilities of an economy. For my analysis, I use the measure of output-side GDP in current PPPs ($CGDP^o$). This variable is divided by the number of persons engaged in employment (emp) to generate a measure of output per worker (gdppc). Similarly, I divide the measure of capital stock in current PPPs (ck) by the number of persons engaged in employment (emp) to generate a measure of each country's capital labour-ratio (k). It should be noted that the capital stock measure in PWT 8.0 differs from earlier versions as well. The new estimates are based on up to six different types of assets which provide a more accurate

[10] See the Appendix for a full list of the variables included in the regressions and their definitions and sources

[11] See Feenstra, Inklaar, and Timmer, 2013 for a discussion of how PWT8.0 differs from earlier versions.

assessment of differences in the composition of capital across countries. Lastly, my regressions include a measure of each country's human capital (hc) stock. This variable is a composite index of human capital based on the average years of schooling from Barro and Lee (2010) and the estimated rates of return for different years of educational attainment estimated by Psacharopolous (1994). All production data reflect 2010 values and are expressed in natural logarithms.

My model controls for a number of geographical variables. Geography can play a key role in economic development as it influence a country's climate, natural endowments (including what crops it can grow), disease burden, and access to knowledge and technology (Diamond, 1999, Gallop et al, 1999; Nordhaus, 2006). Recent research indicates that higher temperatures reduce economic growth in poor countries (Dell, Jones, and Olken, 2012) and can lead to higher levels of political conflict (Hsiang, Burke, and Miguel, 2013). Precipitation appears to have a smaller effect on economic growth although it affects both poor and rich countries (Dell, Jones, and Olken, 2012). In my regressions, I control for both the level of precipitation (in mm) and the average temperature (in Celsius) for each country.

In addition, my regressions control for tropical status (i.e., whether the country has a latitude which lies between 23.5 degrees N and 23.5 degrees S). Several researchers have pointed out that countries in the tropics tend to have lower levels of development than those located in more temperate regions (Easterly and Levine, 2003; Gallop, Sachs, and Mellinger, 1999; Gallop and Sachs, 2001a, 2001b; Sachs, 2003; Bonds et al, 2012). For example, each of the twenty poorest countries in the world in 2010 was located in a tropical climate. By contrast, 18 of the 20 richest countries were located in temperate climates—the two exceptions being Singapore and Brunei. Acemoglu, Johnson, and Robinson (2001) argue that countries in the tropics tend to have weaker institutions than those located in temperate zones because tropical countries did not attract the same level of European settlement during the colonial period as those with temperate climates. Another potential explanation for the poor economic performance of tropical countries is that they have a greater prevalence of vector-borne and parasitic diseases (VBPDs). As explained by Bonds et al (2012): "VBPDs continue to be among the leading causes of morbidity and mortality of poor populations. Unlike directly transmitted diseases, VBPDs spend much of their life cycle outside of the human host¸ in other host species or in free-living stages, and are

thus especially dependent on external conditions" (p. 1). Among VBPDs, malaria claims the most lives in poor countries and results in the largest numbers of DALYs lost per year (Institute for Health Metrics and Evaluation, 2013).

A country's location also matters for access to international markets and ports. Landlocked countries face higher transportation costs and can be at an economic disadvantage when surrounded by poor countries (Collier, 2007). Therefore, all regressions include a dummy variable indicating whether the country is landlocked or not. Similarly, countries that are resource-abundant and rely heavily on mineral or energy exports may be at an economic disadvantage due to "dutch disease" and rent-seeking behavior. There is now a large literature which documents several channels through which abundant natural resources can lead to lower economic growth (see Frenkel, 2010 for a survey of this literature). A country's ability to extract and export such natural resources, however, is likely to be endogenous. To control for this possibility, I include a dummy variable which indicates whether the country has oil reserves or not. This variable is preferred to one which reflects current (or past) values of oil revenues which is likely to be endogenous.[12]

Since the focus of the paper is on the economic impact of malaria, I estimate my model using several different measures of malaria. They are:

1) PAR10 is the proportion of the population living in areas of stable risk of *P. falciparum* transmission in 2010 as reported by Gething et al, 2011. Stable risk areas are defined as those where there is more than 1 reported case of malaria per 1000 individuals each year. *P. falciparum* is the strain of malaria responsible for the largest number of deaths worldwide each year. It is more common in Sub-Saharan Africa than in other regions where less virulent strains are more common (e.g., *P. vivax, P. malariae, and P. ovale*).

2) PAR94 is the proportion of the population living in areas at risk of malaria transmission in 1994 as reported by Sachs (2003). Risk is based on the possibility of exposure to

[12] In earlier specifications, I also included dummy variables which control for the existence of coal reserves and minerals. The estimated coefficients on these variables, however, were never significant and their exclusion did not affect my results. Therefore, I omitted these variables from the final set of regressions.

several strains of malaria—not just *P. falciparum*. Data on the global distribution of *P. falciparum* did not exist when this variable was constructed. While this variable is not directly comparable to PAR10, it provides a useful measure of the potential lagged effects of malaria risk.

3) LnDALYs is the logarithm of the number of disability-adjusted-life-years (DALYs) lost per person due to malaria in 2010. It is constructed by dividing the total number of DALYS lost from malaria in each country by its population. The estimated number of DALYs is obtained from the Institute for Health Metrics and Evaluation (2013). According to these estimates, malaria is the seventh leading cause of DALYs lost worldwide.

4) The Malaria Ecology (ME) index was introduced by Kiszewski et al in 2004. It is based on a country's environmental suitability for supporting the transmission of different malaria vectors. Since the ME index varies with a country's climate (and not its actual cases of malaria), several economists have argued that it represents a good exogenous source of variation for malaria incidence.

A number of studies have used the ME index as an instrumental variable for malaria incidence (Sachs, 2003; Carstensen and Grundlach, 2006; Bhattacharya, 2009). The main weakness of the ME index, however, is that it may be correlated to other factors (e.g., institutions) which affect economic growth. For this study, I do not use the ME index as an instrumental variable in my core specifications. Instead, I rely on the variation in genetic mutations in the G6PD gene as an exogenous source of variation. Unless evolutionary forces which took place thousands of years ago are correlated to factors which affected institutional development, it is unlikely that rates of G6PD deficiency are correlated with differences in institutional quality across countries.

5. Estimation Results

5.1. OLS Regressions

To begin, equation (1) is estimated by Ordinary Least Squares (OLS) using each of the three direct measures of malaria. The results of these regressions are reported in Table 1. The point estimate of β reflects the average change in log output per worker associated with either a one unit increase in malaria prevalence (for PAR10 and PAR94) or a 1% increase in malaria morbidity (for LnDALYs), holding all other variables constant. None of the point estimates of β are significant in any of the OLS regressions, indicating that malaria does not have a significant impact on output per worker. By contrast, the estimated coefficient on "years malaria free" is positive and significant in each specification. The magnitude of this coefficient is 0.003 which indicates that, on average, output per worker increases by 0.3 % per year following malaria elimination, after controlling for the standard production and geography variables. While this number is small in magnitude, its cumulative effect becomes quantitatively important over time. For example, the level of output per worker in a country that eradicated *P. falciparum* 100 years ago is now 35% higher after eliminating this deadly strain of malaria.

Table 1 also reports the point estimates of the input variables which are included in the regression as covariates. The point estimate of the log of (K/L) is significant and quantitatively important. Surprisingly, the point estimate of the log of human capital is insignificant. Its lack of significance suggests that malaria may be correlated with other variables that affect human capital accumulation. The OLS estimates reported in Table 1, however, may be misleading for a number of reasons including omitted variable bias, reverse causality, and measurement error. To tackle these problems, I employ a two-stage least squares estimation which uses rates of G6PD deficiency as an instrumental variable for malaria.

5.2. 2SLS Results

Table 2 reports the 2SLS results where rates of G6PD deficiency are used as an instrumental variable for malaria. The results indicate that greater levels of malaria risk and malaria-related morbidity have a negative impact on economic performance. When rates of G6PD deficiency are used an instrument for the population at stable risk of contracting *P. falciparum*, the β estimate is both significant and quantitatively important. The estimated coefficient of PAR10 is -0.518 (column 2) which indicates that a 1% increase in the stable risk of contracting *P. falciparum* is associated with an average fall of 0.5% in output per worker, holding the standard production and geography variables constant. The β estimate is significant at the 1% level. The reported

confidence intervals contain the unknown true parameters of β with a 95% confidence level. Therefore, the true coefficient of PAR10 is estimated to be between -0.12 and -0.92. As reported in column (5), the size of the estimated coefficient of PAR10 does not change in any meaningful way when regional controls are added. In addition, the β estimate remains significant at the 5% level.

Current levels of malaria prevalence, however, may not capture the full impact of malaria on aggregate income if there are lagged effects of being exposed to malaria. Previous research reveals that malaria eradication increases the level of educational attainment among the young (Bleakley, 2010; Lucas, 2010; Change et al, 2011) and reduces the likelihood of contracting chronic diseases later in life (Hong, 2013). The size of the estimated coefficient of PAR94 indicates that a 1% rise in malaria risk is associated with a 0.5% fall in output per worker (column 3) which is the same as that estimated for PAR10 (column 2). PAR94 is significant at the 5% level. Interestingly, the magnitude of estimated coefficient on PAR94 is much smaller than that estimated by Sachs (2003) when he uses the malaria ecology (ME) index and the share of a country's population in temperate ecozones as instruments. Sachs's β estimates range in value from -1.07 to -1.43. His specification, however, differs considerably from equation (1). Sachs's model controls for institutional quality but it does not control for either human capital or the capital-labour ratio. Nor does he include important geography variables such as tropics, rainfall, temperature, and regional controls which are known to affect economic growth.

Table 2 also reports the results of the estimated effect of ln(DALYs) on the log of output per worker. In column (1) the β estimate of Ln(DALYs) is -0.061 which indicates that a 1% rise in Ln(DALYs) lost from malaria is associated with a 0.06% decline in output per worker. The β estimate is significant at the 1% level when the specification does not control for regional fixed effects. Once again, its significance falls to 5% when regional controls are added. The β estimate reported in column (1) is smaller than that estimated by other studies. Bonds et al (2012) report a β coefficient of -0.40 when they measure the impact of Ln(DALYs) on GDP per capita. Their measure of Ln(DALYs), however, captures a wider range of vector borne and parasitic disease

(of which malaria is the most important) and they do not calculate DALYs on a per capita basis[13].

The relevance of my results, however, depends upon the validity of my estimation strategy. Conventional asymptotic theory no longer holds if the instrumental variable is only weakly correlated with the endogenous explanatory variables. Given a set of weak instruments, any statements about statistical significance and inference may be misleading. Therefore the relevance of the instrumental variable—G6PD deficiency rates—has to be checked. The first-stage regressions provide valuable information about the strength of the instrument. The F-statistics are all large (ranging from 25.1 to 71.63) and highly significant with reported p-values of zero. In addition, the partial R-squared is well above zero for each specification. Somewhat weaker support for my identification strategy is found when the Hausman statistical test is carried out. While the null of exogeneity can be rejected at either the 5% or 10% level when the model is estimated without regional controls (see columns 1-3), the p-values are slightly larger when regional controls are added (see column 4-6).

5.3. Robustness Checks

The exclusion restriction implied by my instrumental variable regression is that rates of G6PD deficiency are correlated with current rates of malaria transmission but not correlated to other characteristics which might affect a country's growth potential. Several robustness checks are carried out to investigate this possibility.

As a first test, I add a measure of the incidence of sickle-cell anaemia to each regression. This variable is defined as the percentage of newborns born with sickle-cell anaemia in each malaria endemic country.[14] The rational for adding this variable is that countries with a high incidence of G6PD deficiency are likely to have a high incidence of sickle cell anaemia as well. As is well known, sickle-cell anaemia is a "loss of function" genetic disorder which causes chronic health problems and often results in early death. If rates of G6PD deficiency are highly correlated to rates of sickle-cell anaemia, it is possible that my instrumental variable is simply picking up the effects of this omitted variable. To address this possibility, I add a measure of

[13] They use the total number of DALYs lost from VBPDs in each country.

[14]The data for sickle-cell anaemia cover all malaria-endemic countries. There are no consistent measures of sickle-cell for malaria-free countries. Since rates of sickle-cell are extremely low in these countries, they are assigned a value of zero.

sickle-cell anaemia to equation (1) and then re-estimate the equation using 2SLS. The results from these regressions are reported in Table 3. As expected, the first-stage regressions indicate that rates of sickle-cell anaemia are positively correlated with each of the malaria variables. The estimated coefficient on sickle cell anaemia is highly significant at the 1% level. The addition of sickle-cell anaemia as a covariate, however, does not change the magnitude of the β estimate in any meaningful way. The significance of the β estimate falls slightly in columns (1-3) to the 5% confidence level. When regional controls are added, its significance falls to the 10% level in columns (4) and (5). The estimated coefficient on PAR94 is not significant in column (6). Overall, the results suggest that malaria and sickle-cell anaemia are capturing different effects—a result which is consistent with current scientific research. According to Hedrick (2011), the natural selection process for G6PD deficiency and sickle-cell anaemia appears to be different. He writes: the "genes are unlinked and are present on different chromosomes" (p. 297).

Another possibility is that malaria and institutional quality are correlated. Since institutional quality is not included in equation (1), the model may suffer from omitted variable bias. Several economists have argued that geography has an indirect—not direct— effect on economic development by influencing how a country's institutions evolve (Engerman and Sokoloff, 2000; Acemoglu, Johnson, and Robinson, 2001, 2002; and Rodrik et al, 2004). While this debate is still on-going, recent research which estimates the effects of both geography and institutional quality on economic performance finds evidence that geographic factors, especially malaria, have an independent effect on economic performance (Sachs, 2003; Cartensen and Gundlach, 2006; Bhattacharya, 2009; Bonds et al, 2012). In this study, I control for institutional quality using a measure of government effectiveness from the World Bank's database on *Worldwide Governance Indicators* (WGI), Version VIII. The following equation is estimated:

$$lny_c = \alpha + \beta \ln k_c + \gamma lnh_c + \delta Malaria_c + \theta Inst_c + \boldsymbol{X_C^G \Pi} + \varepsilon_{\boldsymbol{c}} \qquad (3)$$

where $Inst_c$ is a measure of institutional quality. Given the potential endogeneity of institutional quality, I instrument using one of two variables. I use either AJR's (2001) logarithm of settler mortality rates (LnMORT) or Jones's (2013) logarithm of governors' salaries (LnSAL). Thus I estimate the following first-stage regressions:

$$Inst_c = \mu + \pi lnMORT_c + \sigma \ln k_c + \varphi lnh_c + \boldsymbol{X_C^G \Pi} + \varepsilon_{Inst} \qquad (4a)$$

or

$$Inst_c = \mu + \pi lnSAL_c + \sigma \ln k_c + \varphi lnh_c + \boldsymbol{X_C^G \Pi} + \varepsilon_{Inst}.$$

(4b)

Once again the exclusion restrictions are that $G6PD_c$ and $Mort\ or\ Sal_c$ do not appear in equation (1).

It is important to note that equation (3) is not my preferred specification because it is difficult to tackle two causal questions at the same time. As pointed out by Angrist and Pischke (2009), it doesn't make sense to think of one endogenous variable as a "control" when examining the effects of the other. Keeping this caveat in mind, the estimation of equation (3) should be viewed as a rough test of the possibility that the model suffers from omitted variable bias. The results from these regressions are reported in Table 4. The results suggest that the malaria indicators and "government effectiveness" are measuring different effects. When "government effectiveness" is added as a covariate, the estimated coefficient on malaria remains significant in each of the core specifications (columns 1-3). The β estimate is not significant when both regional controls and a measure of institutional quality are included in equation (3). It is likely that the reduced significance of the β estimates is partially explained by the relatively small sample sizes reported in Table 4. Importantly, the estimated coefficient on "government effectiveness" is not significant in any of the regressions.

The last statistical test performed is an overidentification test which adds one additional instrumental variable to the first-stage of the IV regression in order to test the regressors for exogeneity. To carry out this test, I add an index of malaria ecology (ME) which is the standard instrument for malaria used by other economists. The ME index was constructed by Kisweski et al (2004) and relates the biological and environmental conditions in each country with those necessary for mosquitoes to survive and transmit different vectors of malaria. According to Kiszewski et al, these conditions do not vary with either public health interventions or economic conditions so the ME index is a valid instrument for malaria. Table 5 reports the results of the tests of overriding restrictions. The reported p-values of the Hansen *J* test statistic indicate that the instruments are appropriately uncorrelated with the disturbance process.

6. Conclusion

Since the early 2000s there has been a renewed interest in the economic effects of malaria. Measuring the impact of malaria on aggregate outcomes has proven difficult, however, because of the potential endogeneity of the disease. This paper proposes a new instrumental variable for malaria incidence which exploits the regional variation in genetic mutations which have arisen through natural selection in order to protect individuals against malaria. The source of identification is the heterogeneity in rates of G6PD deficiency across countries. It is argued that rates of G6PD deficiency are an appropriate instrument for malaria incidence because the evolutionary forces which caused the gene mutation were most prominent in countries with a history of malaria transmission. Moreover, G6PD deficiency is asymptomatic in the vast majority of individuals who inherit the blood disorder. Using a two-stage least squares estimation procedure, I find evidence that countries with a higher incidence of malaria (and malaria-related morbidity) have lower output per worker than countries where malaria no longer poses a health risk. Specifically, my results indicate, if *P. falciparum* were totally eradicated (that is, the population at stable risk of *P. falciparum* were reduced to zero), the average level of output per worker in malaria endemic countries would rise by about 50% after controlling for the standard production and geography variables.

To confirm these results, I conduct several robustness checks. The results from these tests support my hypothesis that higher levels of malaria are causally linked to lower levels of output per worker. In addition, the study takes advantage of recent research on the human genome which links diverse genetic adaptations to malaria. My results indicate that past genetic mutations are an appropriate instrumental variable for current levels of malaria incidence. This identification strategy is likely to be appropriate for other diseases as well and it is hoped that future research will build upon these results.

APPENDIX: Definitions and Sources of Variables

Variable	Definition	Source
gdppc	Output-side GDP at current PPPs (2010) divided by number of persons engaged in employment; measured in Millions 2005 US$.	Feenstra, Robert, C., Robert Inklaar, and Marcel Timmer. 2013. PWT Version 8.
K	Capital stock at current PPPs (2010) divided by number of persons engaged in employment; measured in Millions 2005 US$.	Feenstra, Robert, C., Robert Inklaar, and Marcel Timmer. 2013. PWT Version 8.
Hc	Human capital index, based on years of schooling (Barro and Lee, 2010) and the estimated rates of return for different years of educational attainment estimated by Psacharopoulos (1994).	Feenstra, Robert, C., Robert Inklaar, and Marcel Timmer. 2013. PWT Version 8.
ethnic	One minus the Herfindahl index of ethnolinguistic group shares. Values range from 0 to 1. Higher values indicate a higher probability that two randomly selected individuals within the population belong to different groups.	Alesina, et al., 2003.
rain	Average precipitation in mm (2008).	World Bank. 2011. *World Development Indicators*. Downloaded from http://data.worldbank.org/data-catalog/world-development-indicators
temp	Average temperature in Celsius (2006).	Dell et al., 2012.
latitude	Distance from the equator as measured by the absolute value of country-specific latitude in degrees.	La Porta et al. 1999.

APPENDIX: Definitions and Sources of Variables

Variable	Definition	Source
tropics	Dummy variable which takes the value of one if the country has latitude less than 0.235; zero otherwise.	La Porta et al. 1999.
Oil	Dummy variable which takes the value of one if the country has oil reserves; zero otherwise.	World Energy Council. 2010.
landlocked	Dummy variable which takes the value of one if the country does not have access to an ocean; zero otherwise.	Brunsden, D, C. Clarke, I.V. Evans and P. Haggett. 2012.
Africa	Dummy variable which takes the value of one if the country is located in Africa; zero otherwise.	Brunsden, D, C. Clarke, I.V. Evans and P. Haggett. 2012.
Asia	Dummy variable which takes the value of one if the country is located in Asia; zero otherwise.	Brunsden, D, C. Clarke, I.V. Evans and P. Haggett. 2012.
South America	Dummy variable which takes the value of one if the country is located in Africa; zero otherwise.	Brunsden, D, C. Clarke, I.V. Evans and P. Haggett. 2012.
PAR10	Proportion of the population that lives at stable risk of *P. falciparum* (2010). Population in 100,000s living in areas of stable *P. falciparum* risk (*Pf*API ≥0.1 per 1,000 people p.a.).	Gething et al, 2011. Downloaded from Malaria Atlas Project at http://www.map.ox.ac.uk/
PAR94	Proportion of the population that lives at risk of malaria transmission in 1994.	Sachs, J. 2003.

APPENDIX: Definitions and Sources of Variables

Variable	Definition	Source
ME	Malaria Ecology Index which is based on a country's environmental suitability for the transmission of malaria.	Kiszewski et al. 2004.
Eradicate	Dummy variable which takes the value of one if the country has history of malaria but no current malaria transmission; zero otherwise.	Tatem et al, 2013.
YrsFree	Number of years that country has been malaria free. Year of malaria eradication obtained from Tatem et al, 2013.	Tatem et al, 2013.
G6PD	Percentage of population with G6PD deficiency in malaria endemic countries.	Howes et al, 2012. Downloaded from the Malaria Atlas Project at http://www.map.ox.ac.uk/
sickle	Percentage of newborns born with sickle-cell anaemia in malaria endemic countries.	Piel et al, 2013. Downloaded from the Malaria Atlas Project at http://www.map.ox.ac.uk/
gov	Captures the quality of public services and the civil service, including its independence from political pressures. Ranges in value from -2.5 to +2.5.	World Bank. 2013. *World Governance Indicators* (WGI). Downloaded from the World Bank at http://info.worldbank.org/governance/wgi/index.aspx#home
mort	Settler mortality rates in colonies	Acemoglu, Daron, Simon Johnson, and James A. Robinson, 2001.
Sal	Colonial Governors' salaries in 1913.	Jones, P. 2013.

References

Acemoglu, Daron, and Simon Johnson. (2007). 'Disease and Development: the Effect of Life Expectancy on Economic Growth,' *Journal of Political Economy* 115(6), pp. 925-985.

Acemoglu, Daron, Johnson, Simon and James A. Robinson. (2001). 'The Colonial Origins of Comparative Development: An Empirical Investigation.' *American Economic Review*, 91(5), pp. 1369-1401.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. (2002). 'Reversal of Fortune: Geography and Development in the Making of the Modern World Income Distribution,' *Quarterly Journal of Economics*, 117(4), pp. 1231-1294.

Alderman, Harold, John Hoddinott, and Bill McKinsey. (2006). 'Long-Term Consequences of Early Childhood Nutrition,' *Oxford Economic Papers* 58(3), pp. 450-474.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kerlat, and Roman Wacziarg. (2003). 'Fractionalization,' *Journal of Economic Growth* 8, pp. 155-194.

Angrist, Joshua D. and Jörn-Steffen Pischke. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton and Oxford: Princeton University Press.

Ashraf, Quamru H., Ashley Lester, and David N. Weil. (2009). 'When Does Improving Health Raise GDP?' *NBER Macroeconomics Annual 2008,* Volume 23, pp. 157-204.

Barreca, Alan. (2010). 'The Long-Term Impact of In Utero and Postnatal Exposure to Malaria,' *Journal of Human Resources* 45(4), pp. 865-892.

Barro, Robert J. and Jong-Wha Lee. (2010). 'A New Data Set of Educational Attainment in the World, 1950-2010,' *NBER Working Paper* no 15902.

Bhattacharyya, Sambit. (2009). 'Root Causes of African Underdevelopment,' *Journal of African Economies* 18(5), pp. 745-780.

Bleakley, Hoyt. (2007). 'Disease and Development: Evidence from Hookworm Eradication in the American South,' *Quarterly Journal of Economics* 122(1), pp. 73-117.

Bleakley, Hoyt. (2010). 'Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure,' *American Economic Journal: Applied Economics* 2, pp. 1-45.

Bloom, David E. David Canning, and Jaypee Sevilla. (2004). 'The Effect of Health on Economic Growth: A Production Function Approach,' *World Development* 32(1), pp. 1-13.

Bonds, Matthew H, Andrew P. Dobson, and Donald C. Keenan. (2012). 'Disease Ecology, Biodiversity, and the Latitudinal Gradient in Income,' *Public Library of Science Biology* 10(12): e1001456.

Brunsden, D, C. Clarke, I.V. Evans and P. Haggett. (2012). *Oxford Essential World Atlas*. Oxford: Oxford University Press.

Cappellini, M.D. and G. Fiorelli. (2008). 'Glucose-6-Phosphate Dehydrogenase Deficiency,' *The Lancet*¸ Jan. 5-Jan. 11, 371, pp. 64-74.

Cartensen, Kai and Erich Gundlach. (2006). 'The Primacy of Institutions Reconsidered: Direct Income Effects of Malaria Prevalence,' *World Bank Economic Review* 20(3), pp. 309-339.

Carter, R. Mendis. (2002). 'Evolutionary and Historical Aspects of the Burden of Malaria,' *Clinical Microbiology Review* 15, pp. 564-594.

Case, Anne and Christina Paxson. (2008). 'Stature and Status: Height, Ability, and Labour Market Outcomes,' *Journal of Political Economy* 116(3), pp. 499-532.

Case, Anne, Darron, Lubostky, and Christina Paxson. (2002). 'Economic Status and Health in Childhood: the Origins of the Gradient,' *American Economic Review* 92(5), pp. 1308-1334.

Chang, Simon, Belton Fleisher, Seonghoon Kim, and Shi-Yung Liu. (2011). 'Long-term Effects of Early Childhood Exposure on Education and Health: Evidence from Colonial Taiwan,' IZA Discussion Paper No. 5526.

Collier, Paul. (2007). *The Bottom Billion*. Oxford: Oxford University Press.

Cutler, David and Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl. (2010). 'Early-life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India,' *American Economic Journal: Applied Economics* 2(2), pp. 72-94.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. (2012). 'Temperature Shocks and Economic Growth: Evidence from the Last Century,' *American Economic Journal: Macroeconomics* 4(3), pp. 66-95.

Diamond, Jared. (1999). *Guns, Germs, and Steel: the Fate of Human Societies*. NY: W.W. Norton & Co.

Dronamraju, Krishna R. (ed.). (2004). *Infectious Disease and Host-Pathogen Evolution*. Cambridge: Cambridge University Press.

Easterly, William and Robert Levine. (2003). 'Tropics, Germs, and Crops: How Endowments Influence Economic Development,' *Journal of Monetary Economics* 50, pp. 3-39.

Engerman, Stanley and Kenneth Sokoloff. (2000). 'History Lessons: Institutions, Factor Endowments, and Paths of Development,' *Journal of Economic Perspectives,* 14(3), pp. 217-232

Feenstra, Robert, C., Robert Inklaar, and Marcel Timmer. (2013). 'The Next Generation of the Penn World Table,' available for download at www.ggdc.net/pwt.

Frenkel, Jeffrey. (2010). 'The Natural Resource Curse: A Survey,' *NBER Working Paper* 15836.

Gething, P.W., A.P. Patil, D.L. Smith, C.A. Guerra, I.R.F. Elyazar, C.L. Johnston, A.J. Tatem, and S.I. Hay, S.I. (2011). 'A New World Malaria Map: *Plasmodium Falciparum* Endemicity in 2010.' *Malaria Journal*, 10, pp. 378.

Gallop, John Luke and Jeffrey D. Sachs. (2001a). 'The Economic Burden of Malaria,' *American Journal of Tropical Medicine and Hygiene* 64(1), pp. 85-96.

Gallop, John Luke and Jeffrey D. Sachs. (2001b). 'Agriculture, Climate, and Technology: Why are the Tropics Falling Behind?' *American Journal of Agricultural Economics* 82, pp. 731-737.

Gallop, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. (1999). 'Geography and Economic Development,' *International Regional Science Review* 22(2), pp. 179-232.

Grossman, Michael. (2000). 'The Human Capital Model,' in *Handbook of Health Economics*, Volume 1A, ed. Anthony J. Cuyler and Joseph P. Newhouse. Amstedam: Elsevier Science B.V.

Hedrick, P.W. (2011). 'Population Genetics of Malaria Resistance in Humans,' *Heriditary* 107, pp. 283-304.

Hong, Sok Chul. (2013). 'Malaria: An Early Indicator of Later Disease and Work Level,' *Journal of Health Economics* 32, pp. 612-632.

Howes, Rosalind A. and Katherine E. Battle, Ari W. Satyagraha, J. Kevin Baird, and Simon I. Hay. (2013). 'G6PD Deficiency: Global Distribution, Genetic Variants and Primaquine Therapy,' in *Advances in Parasitology*, Volume 81.

Howes, Rosalind E, and Frédéric B. Piel, Anand P. Patil, Oscar A. Nyangiri, Peter W. Gething, Mewayhu Dewi, Mariana M. Hogg, Katherine E. Battle, Carmencita D. Padilla, J. Kevin Baird, and Simon I. Hay. (2012). 'G6PD Deficiency Prevalence and Estimates of Affected Populations in Malaria-Endemic Countries: A Geostatistical Model-Based Map,' *Public Library of Science Medicine* 9(11): e1001339.

Hsiang, Solomon M., Marshall Burke and Edward Miguel. (2013). 'Quantifying the Influence of Climate on Human Conflict,' *Science*, 10.1126/science 1235367.

Institute for Health Metrics and Evaluation. (2013). *The Global Burden of Disease: Generating Evidence, Guiding Policy*. Seattle WA: IHME.

Johnson, MK, Clark, TD, Njama-Meya, D, Rosenthal, PJ, and Parik. (2009). 'Impact of the Method of G6PD Deficiency Assessment on Genetic Association Studies of Malaria Susceptibility,' *Public Library of Science Genetics* 6(6): e1000924.

Jones, Patricia. (2013). 'History Matters: New Evidence on the Long-run Impact of Colonial Rule on Institutions,' *Journal of Comparative Economics* 41, pp. 181-200.

Kiszewski, Anthony, Andrew Mellinger, Andrew Spellman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs. (2004). 'A Global Index Representing the Stability of Malaria Transmission,' *American Journal of Tropical Medicine and Hygiene* 705(4), pp. 486-498.

Kwiatkoski, Dominic P. (2005). 'How Malaria Has Affected the Human Genome and What Human Genetics can teach us about Malaria,' *American Journal of Human Genetics* 77, pp. 171-190.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. (1999). 'The Quality of Government,' *Journal of Law, Economics, and Organization* 15, pp. 445-470.

Lucas, Adrienne M. (2010). 'Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka,' *American Economic Journal: Applied Economics* 2(2), pp. 46-71.

Maluccio, John A., John Hoddinot, , Jere R. Behrman, Reynaldo Matorrell, Agnes Quisumbing, and Aryeh D. Stein. (2009). 'The Impact of Improving Nutrition during Early Childhood on Education among Guatemalan Adults,' *Economic Journal* 119, pp. 734-763.

Miguel, Edward and Michael Kremer. (2004). 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,' *Econometrica* 72(1), pp. 159-217.

Montalvo, Jose G. and Marta Reynal-Querol. (2007). 'Fighting Against Malaria: Prevention Wars While Waiting for the 'Miraculous' Vaccine,' *The Review of Economics and Statistics* 89(1), pp. 165-177.

Nkhoma, Ella T. and Charles Poole, Vani Vannappagari, Susan A. Hall, and Ernest Butler. (2009). 'The Global Prevalence of Glucose-6-Phosphate Dehydrogenase Deficiency: A Systematic Review and Meta-Analysis,' *Blood Cells, Molecules, and Diseases* 42, pp. 267-278.

Nordhaus, William D. (2006). 'Geography and Macroeconomics: New Data and New Findings,' *Proceedings of the National Academy of Science* 103(10), pp. 3510-17.

Packard Randall M. (2009). 'Roll Bank Malaria, Roll in Development? Reassessing the Economic Burden of Malaria,' *Population and Development Review* 35(1), pp. 53-87.

Peters, Anna L. and Cornelis J.F. Van Noorder. (2009). 'Glucose-6-phospate Dehydrogenase Deficiency and Malaria: Cytochemical Detection of Heterozygous G6PD Deficiency in Women,' *Journal of Hystochemistry and Cytochemistry* 57(11), pp. 1003-1011.

Piel F.B., A.P. Patil, R.E. Howes, O.A. Nyangiri, P.W. Gething, M. Dewi, W.H. Temperley, T.N. Williams, D.J. Weatherall, and S. I. Hay. (2012). 'Global epidemiology of sickle haemoglobin in newborns: a contemporary geostatistical model-based map,' *The Lancet*, doi:10.1016/S0140-6736(12)61229-X.

Psacharopoulos, George. (1994). 'Returns to Investment in Education: A Global Update,' *World Development* 22(9), pp. 1325-1343.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. (2004). 'Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development,' *Journal of Economic Growth,* 9, pp. 131-165.

Ruwende, C., Khoo, SC, Snow, RW, Yates, SNR, Kwiatowski, D. 1995. "Natural Selection of hemi- and heterozygotes for G6PD deficiency in Africa by Resistance to Severe Malaria," *Nature* 376: 24-249.

Sachs, Jeffrey D. (2003). 'Institutions Don't Rule: Direct Effects of Geography on Per Capita Income,' *NBER Working Paper* 9490.

Sachs, Jeffrey D. and Pia Malaney. (2002). 'The Economic and Social Burden of Malaria,' *Nature* 415(7), pp. 680-685.

Schultz, T. Paul. (2002). 'Wage Gains Associated with Health as a Form of Health Human Capital,' *A.E.R. Papers and Proceedings* 92 (May), 92(2), pp. 349-53.

Steckel, Richard H. (1995). 'Stature and Standard of Living,' *Journal of Economic Literature*, 33(4), pp. 1903-19040

Strauss, John and Duncan Thomas. (1997). 'Health and Wages: Evidence on Men and Women in Urban Brazil,' *Journal of Econometrics* 77, pp. 159-185.

Tatem, Andew J, Peter W. Gething, David L. Smith, and Simon I. Hay. (2013). 'Urbanization and the Global Malaria Recession,' *Malaria Journal* 12:133.

Tishkoff, Sarah A., Robert Varkonyi, Nelie Cahinhinan, Salem Abbes, George Argyropoulos, Giovanni Destro-Bisol, Anthi Drousiotou, Bruce Dangerfield, Gerard Lefranc, Jacques Loiselet, Anna Piro, Mark Stoneking, Antonio Tagarelli, Giuseppe Tagarelli, Elias H, Touma, Scott M. Williams, and Andrew G. Clark. (2001). 'Haplotype Diversity and Linkage Disequilibrium at Human G6PD: Recent Origin of Alleles that Confer Malaria Resistance,' *Science* 293(5529), pp. 455-462.

Venkatataramani, A. (2012). 'Early Life Exposure to Malaria and Cognition in Adulthood: Evidence from Mexico,' *Journal of Health Economics* 31(5), pp. 767-780.

Weatherall, D.J. (2008). 'Generic Variation and Susceptibility to Infection: the Red Cell and Malaria,' *British Journal of Haemotology* 141, pp. 276-286.

Weil, David N. (2007). 'Accounting for the Effect of Health on Economic Growth,' *Quarterly Journal of Economics* 122(3), pp. 1265-1306.

World Bank. (2013). *World Development Indicators*, downloaded from http://data.worldbank.org/data-catalog/world-development-indicators.

World Bank. (2013). *World Governance Indicators*, downloaded from http://data.worldbank.org/data-catalog/worldwide-governance-indicators.

World Energy Council (2010) *Survey of Energy Resources, 2010* (London: World Energy Council).

World Health Organization. (2012). *World Malaria Report 2012*. World Health Organization: Geneva.

World Health Organization. (2013). *Malaria Fact Sheet*. Downloaded from http://www.who.int/mediacentre/factsheets/fs094/en/index.html.

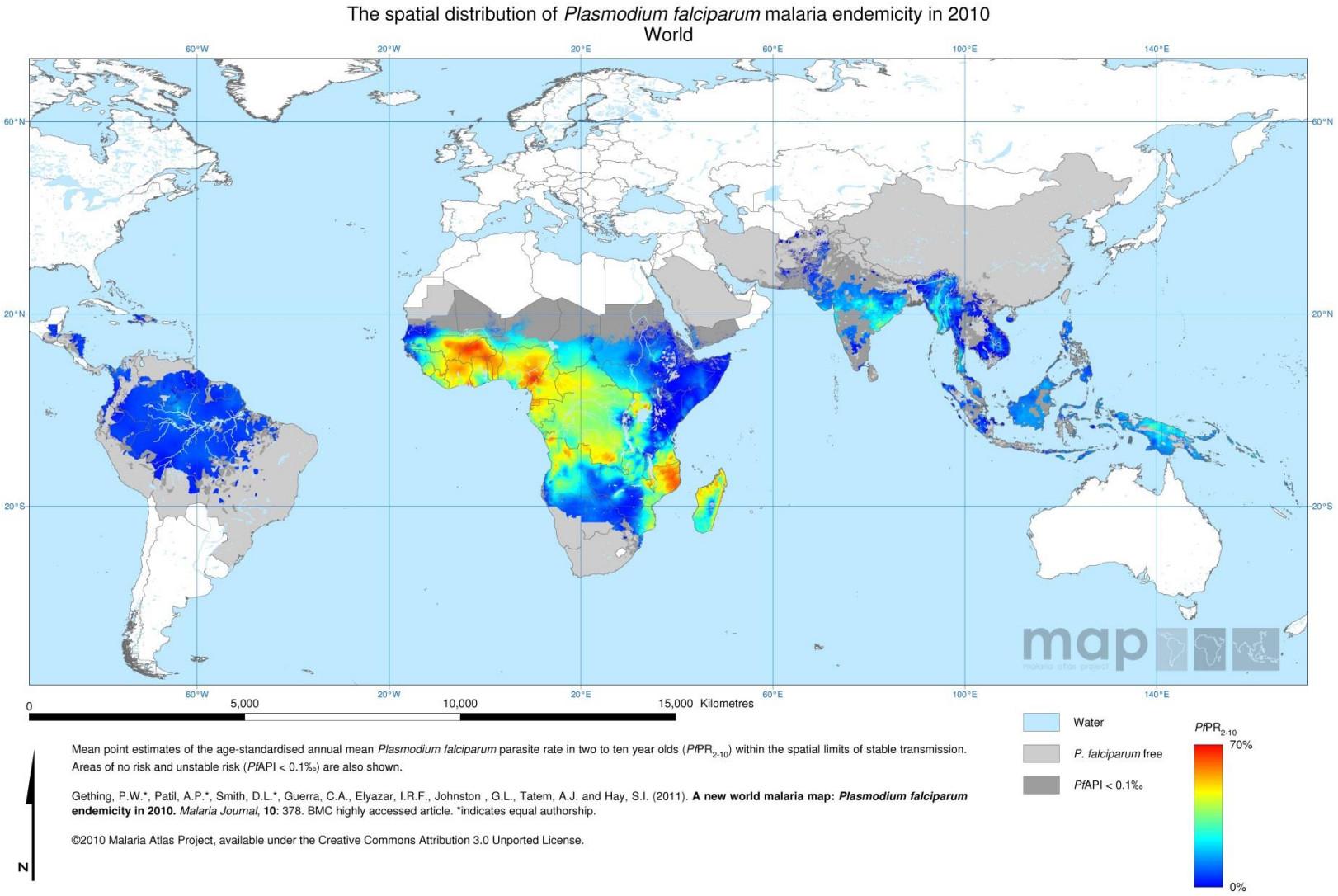

Fig. 1*: Global Distribution of P. Falciparum*
Source: Gething et al, 2011

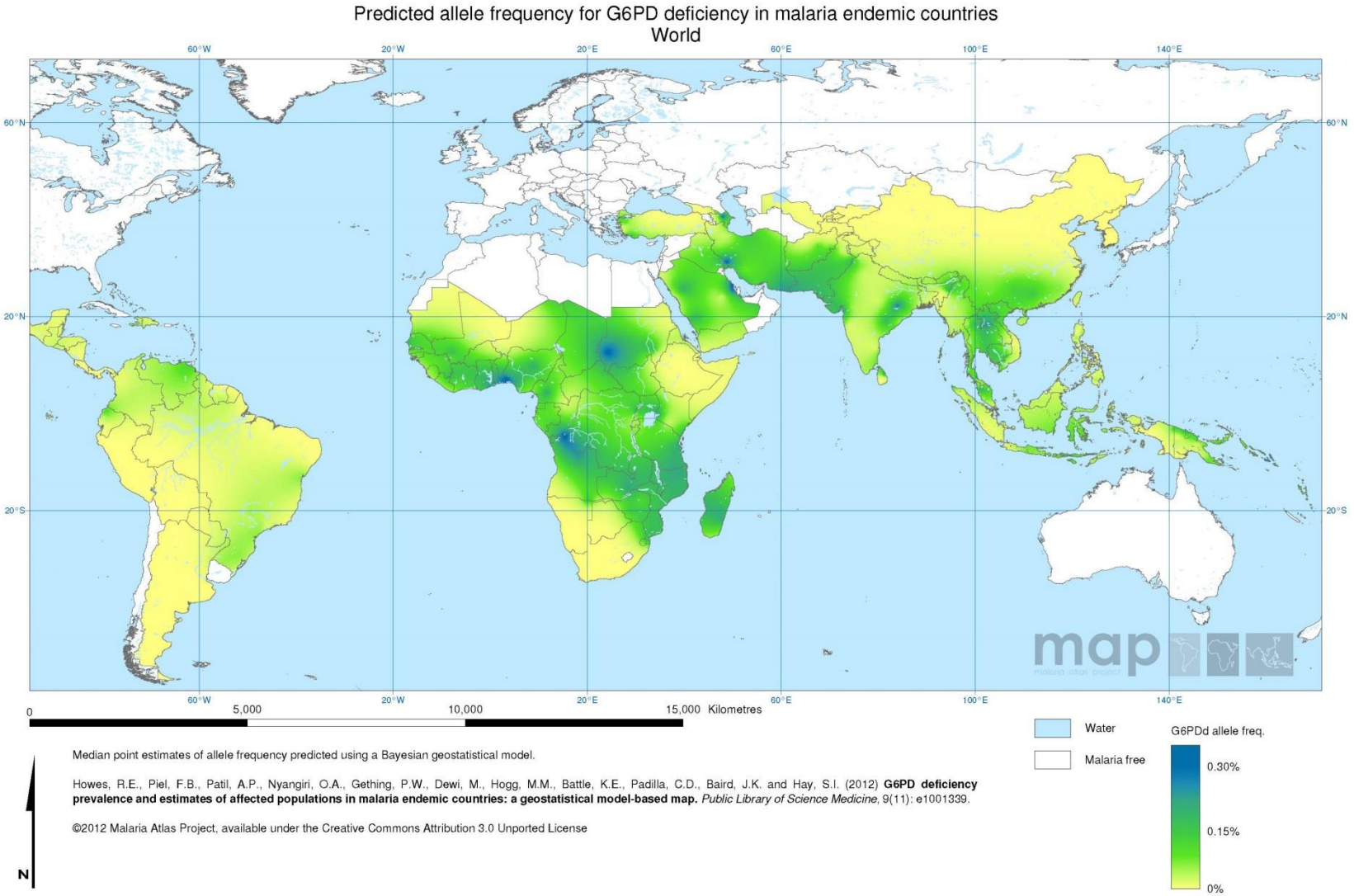

Fig. 2*: Global Distribution of G6PD*
Source: Howes et al, 2012

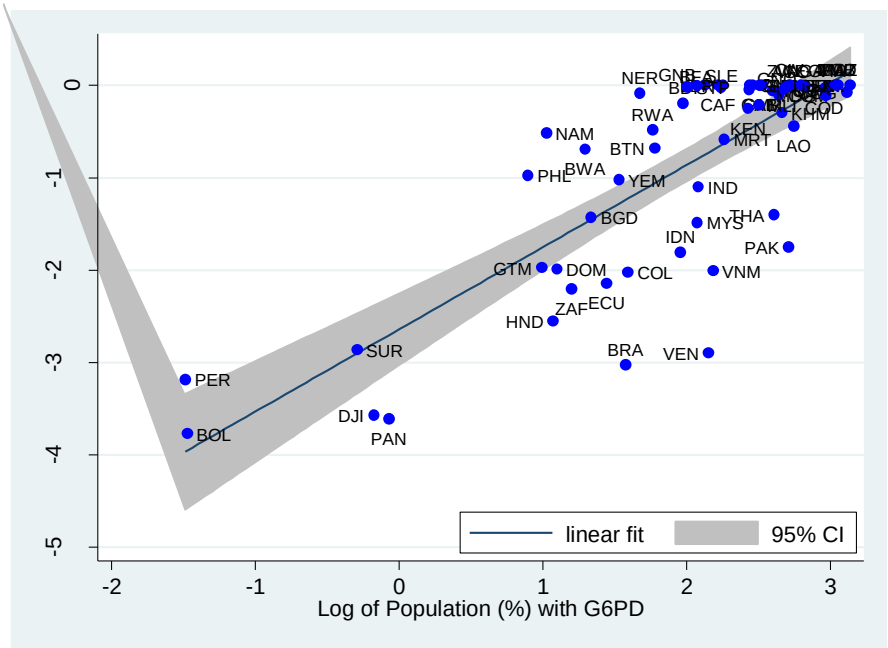

Fig. 3*: Malaria Incidence and G6PD Deficiency*

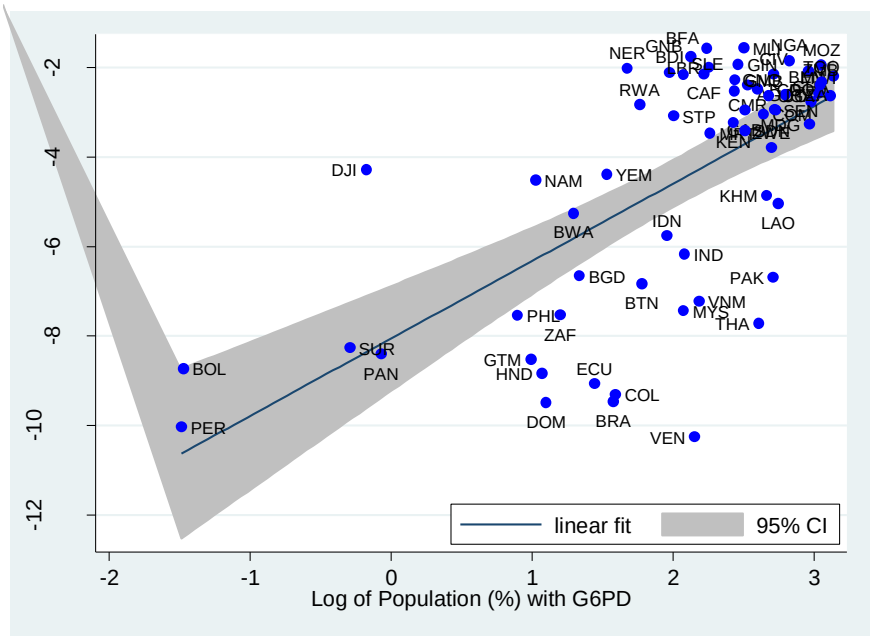

Fig. 4*: Ln(DALYs) and G6PD Deficiency*

Table 1
OLS Estimates: Malaria and Economic Development

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(DALYs)	PAR10	PAR94	Ln(DALYs)	PAR10	PAR94
Malaria Measure	-0.019	-0.209	-0.132	-0.016	-0.188	-0.117
	(0.016)	(0.161)	(0.177)	(0.017)	(0.167)	(0.187)
Ln(K/L)	0.782***	0.786***	0.778***	0.786***	0.782***	0.770***
	(0.060)	(0.060)	(0.063)	(0.063)	(0.064)	(0.070)
Ln(Human Capital)	0.457	0.442	0.407	0.375	0.367	0.28
	(0.291)	(0.286)	(0.324)	(0.272)	(0.273)	(0.285)
Past Malaria Dummy	-0.143	-0.143	-0.109	-0.174	-0.145	-0.147
	(0.100)	(0.101)	(0.098)	(0.108)	(0.101)	(0.100)
Years Malaria Free	0.003**	0.003**	0.003**	0.003***	0.003***	0.003***
	(0.099)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Regional Controls	No	No	No	Yes	Yes	Yes
R-squared	0.95	0.95	0.95	0.95	0.95	0.95
Observations	119	119	116	119	119	116

Notes: The table reports OLS estimates with robust standard errors. The dependent variable is the natural log of GDP per worker in 2010 at chained PPPs. PAR10 indicates the proportion of the population at risk of *P. falciparum in 2010*. PAR94 indicates the proportion of the population at risk of malaria (all strains) in 1994. Ln(DALYs) reflects the average number of years of a healthy life lost due to malaria in 2010. All regressions include controls for average temperature, average precipitation, level of ethnic fractionalization, and dummy variables for tropics, landlocked, and oil. The regional controls are dummy variables for Africa, Asia, and South America. Coefficients are reported with standard errors in parentheses. *, **, and *** indicate significance levels at the 10%, 5%, and 1% levels, respectively.

Table 2
2SLS and Reduced Form Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: First-stage 2SLS Estimates			
Malaria Measure	Ln(DALYs)	PAR10	PAR94	Ln(DALYs)	PAR10	PAR94
G6PD (% of Population)	0.277***	0.033***	0.029***	0.228***	0.027***	0.026***
	(0.046)	(0.004)	(0.005)	(0.045)	(0.004)	(0.005)
F-statistic	36.86	71.34	38.87	25.09	61.07	28.68
p-value	0.00	0.00	0.00	0.00	0.00	0.00
Partial R-squared	0.26	0.40	0.27	0.20	0.37	0.22
Hausman test (p-value)	0.05	0.06	0.05	0.07	0.07	0.11
			Panel B: Second-stage 2SLS Estimates			
Malaria Measure	-0.061***	-0.518***	-0.534**	-0.068**	-0.561**	-0.504*
	(0.025)	(0.204)	(0.235)	(0.031)	(0.251)	(0.266)
Ln(K/L)	0.774***	0.765***	0.746***	0.787***	0.775***	0.752***
	(0.046)	(0.046)	(0.049)	(0.046)	(0.045)	(0.047)
Ln(human capital)	0.270	0.302	0.205	0.219	0.260	0.108
	(0.258)	(0.246)	(0.263)	(0.254)	(0.238)	(0.261)
Eradicated Malaria	-0.230**	-0.112	-0.134	-0.301**	-0.162	-0.188*
	(0.118)	(0.104)	(0.106)	(0.132)	(0.104)	(0.107)
Years Malaria Free	0.003**	0.004**	0.003**	0.004**	0.004***	0.004***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Regional Controls	No	No	No	Yes	Yes	Yes
Observations	119	119	116	119	119	116

Notes: Dependent variable is the natural log of GDP per capita in 2010 at chained PPPs. PAR10 indicates the proportion of the population at risk of *P. falciparum* in 2010. PAR94 indicates the proportion of the population at risk of malaria (all strains) in 1994. Ln(DALYs) reflects the average number of years of a healthy life lost due to malaria in 2010. All regressions include controls for average temperature, average precipitation, level of ethnic fractionalization, and dummy variables for tropical location, being landlocked, and having oil reserves. The regional controls are dummy variables for Africa, Asia, and South America. Coefficients are reported with standard errors in parentheses. *, **, and *** indicate significance levels at the 10%, 5%, and 1% levels, respectively.

Table 3
Robustness Check: Addition of Sickle-Cell as Covariate

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: First-stage 2SLS Estimates			
Malaria Measure	Ln(DALYs)	PAR10	PAR94	Ln(DALYs)	PAR10	PAR94
G6PD (% of Population)	0.242***	0.026***	0.024***	0.217***	0.024***	0.023***
	(0.049)	(0.004)	(0.005)	(0.047)	(0.004)	(0.005)
Sickle Cell (% of Newborns)	0.107*	0.022***	0.017***	0.026	0.016***	0.017***
	(0.056)	(0.004)	(0.006)	(0.06)	(0.004)	(0.006)
F-statistic	24.78	46.08	25.22	21.14	45.32	20.96
p-value	0.00	0.00	0.00	0.00	0.00	0.00
Partial R-squared	0.19	0.31	0.20	0.17	0.31	0.17
Hausman test (p-value)	0.09	0.06	0.06	0.15	0.09	0.14
			Panel B: Second-stage 2SLS Estimates			
Malaria Measure	-0.059**	-0.560**	-0.587*	-0.057*	-0.525*	-0.518
	(0.031)	(0.284)	(0.305)	(0.034)	(0.304)	(0.323)
Sickle Cell (% of Newborns)	-0.002	0.004	0.005	-0.001	-0.004	0.002
	(0.010)	(0.012)	(0.011)	(0.009)	(0.011)	(0.012)
Ln(K/L)	0.775***	0.762***	0.741***	0.792***	0.778***	0.749***
	(0.047)	(0.048)	(0.052)	(0.045)	(0.046)	(0.050)
Ln(human capital)	0.276	0.292	0.183	0.247	0.268	0.105
	(0.262)	(0.254)	(0.279)	(0.249)	(0.241)	(0.273)
Eradicated Malaria	-0.228*	-0.107	-0.130	-0.292	-0.168	-0.187*
	(0.020)	(0.105)	(0.107)	(0.129)	(0.104)	(0.106)
Years Malaria Free	0.004**	0.004**	0.003**	0.004***	0.004***	0.004***
	(0.010)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Regional Controls	No	No	No	Yes	Yes	Yes
Observations	119	119	116	119	119	116

Notes: Same as Table 3.

Table 4

Robustness Check: Addition of Institutional Quality Measure

	(1)	(2)	(3)	(4)	(5)	(6)
	Uses LnMORT as IV for Institutions			*Uses LnSAL as IV for Institutions*		
Panel A: First-stage 2SLS Estimates						
Malaria Measure/Govt Effectiveness	Ln(DALYs)	PAR10	PAR94	Ln(DALYs)	PAR10	PAR94
G6PD (% of Population)	0.348***	0.030***	0.038***	0.161***	0.025***	0.012**
	(0.081)	(0.008)	(0.010)	(0.048)	(0.006)	(0.006)
LnMort	-0.150	-0.150	-0.158	0.449***	0.449***	0.452***
	(0.103)	(0.103)	(0.103)	(0.121)	(0.121)	(0.118)
Panel B: Second-stage 2SLS Estimates						
Malaria Measure	-0.079**	-0.872*	-0.609*	-0.085	-0.556	-1.233
	(0.039)	(0.499)	(0.327)	(0.058)	(0.352)	(1.010)
Government Effectiveness	0.008	-0.421	-0.140	0.237	0.115	-0.223
	(0.473)	(0.473)	(0.437)	(0.203)	(0.179)	(0.404)
Ln(K/L)	0.867***	0.912***	0.817***	0.518***	0.545***	0.613***
	(0.122)	(0.141)	(0.111)	(0.122)	(0.114)	(0.174)
Ln(human capital)	0.031	0.504	0.114	1.132*	1.177*	1.174
	(0.668)	(0.719)	(0.574)	(0.526)	(0.469)	(0.568)
Eradicated Malaria	-0.208	0.213	-0.009	-0.430	0.109	-0.294
	(0.368)	(0.387)	(0.318)	(0.566)	(0.242)	(0.555)
Years Malaria Free	0.006	0.008	0.006	0.003	0.002	0.004
	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)	(0.007)
Regional Controls	No	No	No	No	No	No
Observations	48	48	47	42	42	40

Notes: LnMort is the logarithm of settler mortality rates from AJR (2001). *LnSAL* is the logarithm of governors' salaries from Jones (2013).

Table 5
Overidentification Tests

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First-stage 2SLS Estimates						
Malaria Measure	Ln(DALYs)	PAR10	PAR94	Ln(DALYs)	PAR10	PAR94
G6PD (% of Population)	0.228***	0.024***	0.026***	0.202***	0.022***	0.024***
	(0.040)	(0.005)	(0.005)	(0.045)	(0.004)	(0.005)
Malaria Ecology Index	0.138***	0.024***	0.008*	0.082*	0.019***	0.009**
	(0.043)	(0.005)	(0.005)	(0.047)	(0.005)	(0.005)
F-statistic	21.85	52.51	14.69	10.67	26.57	12.41
p-value	0.00	0.00	0.00	0.00	0.00	0.00
Partial R-squared	0.32	0.62	0.29	0.22	0.51	0.25
Panel B: Second-stage 2SLS Estimates						
Malaria Measure	-0.052**	-0.383**	-0.538**	-0.058**	-0.422**	-0.508**
	(0.024)	(0.179)	(0.263)	(0.028)	(0.199)	(0.251)
Ln(K/L)	0.768***	0.763***	0.748***	0.779***	0.768***	0.751***
	(0.060)	(0.058)	(0.070)	(0.065)	(0.061)	(0.073)
Ln(human capital)	0.253	0.319	0.186	0.159	0.230	0.105
	(0.286)	(0.271)	(0.293)	(0.272)	(0.253)	(0.280)
Eradicated Malaria	-0.202**	-0.099	-0.136	-0.276**	-0.151	-0.190*
	(0.096)	(0.094)	(0.094)	(0.116)	(0.095)	(0.102)
Years Malaria Free	0.003**	0.004***	0.003**	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hansen J (p-value)	0.80	0.52	0.88	0.95	0.71	0.94
Regional Controls	No	No	No	Yes	Yes	Yes
Observations	117	117	116	117	117	116

Notes: Same as Table 3.