Bocconi

WORKING PAPER SERIES

ISSN 2785-1079

Number 10

SEPTEMBER 2020

VECTOR-BORNE DISEASES AND ECONOMIC ACTIVITY: EVIDENCE FROM HISTORICAL FARMER PRODUCTIVITY IN THE US

Maurizio Malpede

Università Bocconi
GREEN
Centre for Geography, Resources, Environment, Energy and Networks

Vector-Borne Diseases and Economic Activity: Evidence from Historical Farmer Productivity in the US

Maurizio Malpede*†

September 17, 2020

Abstract

This study provides an estimation of the causal relationship between the reduction in malaria transmission and farmer agricultural productivity. Exploiting exogenous geographic variations in the stability of malaria and using historical disaggregated county data for the US together with a robust quasi-experimental approach, I show that the eradication of malaria led to approximately one fifth of the farmer agricultural productivity growth in the US. Using historical spatial data on cropland distribution within the US, I also show that the positive effect was entirely due to better health conditions rather than a greater availability of arable land. No effect is found on agricultural output per capita for more endemic counties, suggesting that the increase in farm output was compensated by the increase in population. Robustness checks from geographic variations in malaria stability within neighboring counties along with placebo treatments reinforce the positive effect of the eradication of malaria in the US on farmer agricultural productivity.

Key words: Malaria, Natural Resources, Agricultural Productivity

JEL Codes: I15, N31, N32, O13, Q12

1 Introduction

Malaria and associated vector-borne diseases are critical concerns in both environmental and development economics. Moreover, future climatic conditions caused by anthropogenic activities will likely be more favorable to the transmission of malaria (Caminade et al., 2014).

*GREEN Centre, Bocconi University. Email: maurizio.malpede@unibocconi.it Latest version: https://sites.google.com/view/mauriziomalpede/research

†I am very grateful for the abundance of support, guidance, and helpful comments from Lucia Corno and Rosario Crino'. I am grateful to Fondazione Romeo ed Enrica Invernizzi and the Laboratory for Effectiveness of Anti-poverty Policies (LEAP, Bocconi University). I also thank all seminar participants at the European Association of Environmental and Resources Economists (EAERE), Collegio Carlo Alberto, Cattolica University and University of Milan-Bicocca for their valuable feedbacks.

As a result, the transmission of vector-borne diseases will worsen in areas of the world which are currently affected by malaria. In addition, malarious weather conditions will likely affect historically low endemic countries such as the UK and central Europe (Medlock and Leach, 2015).

While the transmission of malaria as well as general health conditions constitute an important determinant in explaining comparative socio-economic development, medical progresses have shown to increase life expectancy and labor productivity (Deaton, 2013; Bleakley et al., 2014). However, the associated increase in population might generate adverse effects on the economic development and political stability (Acemoglu and Johnson, 2007; Acemoglu et al., 2020).

Therefore, understanding the historical socio-economic impacts of malaria both on the availability of natural resources and on population is key to predict how economies will be economically affected by climate change in the future.

This study provides evidence of the historical impact of the eradication of malaria on farmer agricultural productivity in the US and investigates the underlying mechanisms.

Results reveal that: first, using exogenous variations in climatic conditions more or less suitable for the transmission of malaria, the eradication campaigns significantly increased both the agricultural output and farmer productivity of counties with higher initial malaria suitability conditions compared to counties with lower initial malarious conditions. Second, I show that the boost in farmer productivity was not due to the greater amount of arable land of more endemic counties due to wetland conversion into arable land. Finally, the availability of an effective cure against malaria also increased population in the subsequent decades, offsetting the boost in agricultural output. As a conclusion, although the eradication of malaria did cause an increase in natural resources and productivity, the amount of agricultural output per capita did not increase in more endemic counties after the eradication of malaria.

The empirical analysis relies on a treatment/control strategy using a difference-in-difference (DID) estimation comparing farmer productivity levels between US counties that had climatic conditions more suitable for the transmission of malaria with counties that were less suitable for a stable transmission of the vector-borne disease. This approach is permitted since sufficient time has passed that we can evaluate the long-term consequences of the eradication. Furthermore, the United States are well-suited for this analysis for two reasons: first, the US show a remarkable degree of heterogeneity in weather variables which in turn allows for a sufficient variability in the malaria stability values; and second, reliable disaggregated data (i.e. at county level) on historical agricultural output and productivity are available for a long time. The identification strategy compares highly malarious counties to less endemic counties within the US. It is, therefore crucial that the identification of malaria endemic counties is exogenous (i.e. it does not depend upon socio-economic conditions which might affect the transmission of the latter). For this reason, I use a spatial time-invariant malaria suitability index created by Kiszewski et al. (2004) and based upon climatic conditions which are more or less suitable for the reproduction of two particular species of mosquitoes, namely: Plasmodium falciparum and Plasmodium vivax (Figure 1).

In addition to the Malaria Stability Index (MSI) this study uses a second measure of endemicity of malaria i.e. the Malaria Endemicity Index (MEI) developed by Hay et al. (2004) who have digitized an old map produced by Lysenko and Semashko (1968) showing historical malaria geography and prevalence. Differently from the MSI, the MEI captures actual distribution of malaria in 1900, just before the onset of vector control polices, as shown in Figure 2. The main difference between the MSI and the MEI is that while the former is obtained using exclusively weather conditions more or less suitable for a stable transmission of malaria throughout the year, the latter relies on actual historical presence of malaria and therefore captures the proportion of people affected by the disease in different areas.

The exogenous variations in suitability of weather conditions for the transmission of malaria allow for the comparison between highly malaria suitable and less malaria suitable counties within the U.S. to clearly determine the effects of the eradication campaigns on farmer agricultural productivity. Results obtained with the MSI are then compared with those obtained with the MEI index. Population, weather and agricultural controls are included in the regressions with the inclusion of year, county fixed effects and state specific linear trends. Furthermore, a series of robustness checks from geographic variations in malaria stability index *within neighboring counties* and placebo treatments are performed to prove the validity of the estimated coefficients.

Results reveal that the eradication campaigns which took place between the 1900s and the 1940s (with the administration of quinine and drainage of wetlands, and the development of new effective drugs and chemical components later) are estimated to have had substantial effects on the historical farmer productivity growth in the US. In particular, a unit increase in the Malaria Stability Index (MSI) is associated with 18 percentage point increase in total county farmer agricultural productivity defined as the county farm output value per farmer[1]. One mechanism through which malaria transmission is tackled is land use conversion. In the U.S. drainage of swamps and wetland is one of the oldest and commonest forms of land modification undertaken to improve health conditions and lower the transmission of vector

[1]Although the stability index developed by Kiszewski et al. (2004) ranges from 0 for malaria free areas to 39 for areas with weather conditions extremely suitable for a persistent reproduction of the Plasmodium falciparum and Plasmodium vivax species, values in the U.S. ranges mainly from 0 to 2 with values greater than 0.06 being defined as slightly malarious

borne diseases such as yellow fever and malaria. For these reasons surface water removal was a predominant public policy objective in United States during the 20^{th} century. This conversion of unused wetland into arable and more productive land might have increased the agricultural output of endemic areas and in turn agricultural productivity. I test this hypothesis. Results show that vector-borne control policies did not increase the amount of arable land of more endemic counties compared to less endemic ones after 1900s.

A second channels through which the transmission of malaria might impact agriculture productivity is linked to the poor health conditions of farmers, who are notably the most affected part of the population. The whole harvest indeed will be negatively affected if a farmer caches malaria during the harvesting. It is therefore reasonable to conclude that more endemic areas were likely to be less productive than non endemic ones. The adopted strategy allows to show whether this is the case.

The results of the analysis lead us to conclude that the increase in farmer productivity of highly malarious counties was entirely due to the increased labor productivity of farmers resulting from better health conditions after the elimination of malaria. On the other hand, the eradication of the vector-borne disease also increased population of more endemic areas (Gooch, 2017). The present analysis also shows that the total agricultural output per capita did not significantly change in more endemic counties after malaria was effectively controlled.

A potential threat to the validity of the empirical analysis comes from the possible endogeneity of the treatment. In other words more endemic counties could have pushed the central government to receive the newly available treatments *before* non endemic ones. To overcome this issue first, the study relies on exogenous malaria suitability index rather than data on infection rates. Second, Bleakley (2007), Cutler et al. (2010) and more recently Gooch (2017) show that the reduction in the burden of malaria clearly resulted from critical scientific innovations coming overwhelmingly from outside the highly endemic counties and which have culminated with better understanding of the origins of the disease and the discover of new drugs and chemical components[2]. This mitigates the usual concern about policy endogeneity. Furthermore, in the regressions I include state specific linear trends to control for changes in farmer productivity resulting from state specific policies. Finally, and most importantly, the empirical strategy compares also neighboring counties belonging to the same state (and therefore subject to the same federal policies) which had a great degree of heterogeneity in malaria stability values[3]. The evidence presented from the neighboring counties analysis reinforce the positive effect of malaria eradication on farmer agricultural

[2]The discovery of quinine followed by the improvements in its usage and the widespread availability of the DDT are examples of scientific progresses which led to the substantial reduction in malaria transmission

[3]I specifically compare those counties highly suitable for malaria with neighboring counties which present weather conditions not suitable for the transmission of the disease.

productivity.
A further possible limitation of the results is constituted by the fact that many endogenous factors affecting farm productivity cannot be considered in this study. Examples of such endogenous factors are government policies and investments. The use of potential additional control variables is limited by the lack of such historical variables or plausible instruments. That said however, the hypothetical inclusion of such control variables would absorb part of the effects of malaria treatment to farmer agricultural productivity because government policies, for instance, are investments in adapting to circumstances, including the effects of malaria and other vector borne diseases. For this reason we would not want to remove such adaptation from the estimates, because they are part of the story. The neighboring counties analysis partially alleviates this issue since only counties belonging to the same state, being neighbors and showing substantially different stability values are considered.

This paper speaks to several literatures. First, understanding the deep causes of divergent economic development has always been fascinating for economists and historians and one documented cause of the latter is represented by health conditions. How different health conditions impact the socio-economic development of a society has been addressed both at the individual level and at the aggregated one. While micro empirical studies have examined the effects of the eradication of malaria on human capital, finding a positive long-term relationship at the individual level (Bleakley, 2003, 2009; Cutler et al., 2010; Percoco, 2013), aggregate macroeconomic effects of disease eradication on the economy are still debated (Acemoglu and Johnson, 2007; Ashraf et al., 2008) and more recently Hansen and Lønstrup (2015) and Gooch (2017).

However, given that malaria is generally more prevalent in rural areas compared to urban ones, with farmers being the most affected by the diseases (Robert et al., 2003; Tatem et al., 2008), the eradication of the vector-borne disease might generate an exogenous positive agricultural productivity shock. This effect has yet to be examined empirically.

The effects of malaria and its eradication during the 20^{th} century has been the object of several economic studies. Specifically, Bleakley (2003, 2009), Cutler et al. (2010) and Percoco (2013) have shown how improved health (due to vaccination and disease prevention programs) affects GDP indirectly through educational gains (better children's health translates to higher education levels and in turn to greater future income). These gains at individual level will directly translate to gain in GDP per capita unless the increase in population (also due to lower mortality) outweighs these productivity and educational gains. On the other hand, Acemoglu and Johnson (2007), Ashraf et al. (2008) and more recently Hansen and Lønstrup (2015) have shown that the decreased mortality rate due to scientific advancements has positively impacted population while negligible (if any) effects were found on GDP per capita.

With particular focus on malaria, Gooch (2017) has shown the positive effect of its eradication campaigns on population and population density during the 20^{th} century. Thus, while micro studies on individual level have shown a positive impact of the eradication of malaria, aggregate effects on the economy are still debated.

Establishing a clear direction in the relationship between malaria and farmer productivity growth is crucial to understand the causes of historical economic development both between countries and within nations. Second, this paper also speaks to the economic literature focusing on the relationship between agricultural productivity and economic growth. Gollin et al. (2014) examined world agricultural productivity rates using new disaggregated data concluding that understanding agricultural productivity is at the heart of understanding world income inequality. Similar conclusions were reached in Bustos et al. (2016) who studied the effects of the adoption of new agricultural technologies on structural transformation in Brazil, finding that exogenous higher agricultural productivity shock caused by technological change led to industrial growth. Therefore, assessing the historical effects of malaria on farmer agricultural productivity within the US could also be crucial to understand differences in the economic development of different US areas.

Third, estimating the relationship between malaria eradication and farmer agricultural productivity is crucial to predict how economies which nowadays rely predominantly on agriculture (i.e. Sub-Saharan African, Latin American and South-East Asian countries) will be affected from government policies aimed at eradicating vector borne diseases. To this regard, climatic conditions particularly suitable for the diffusion of vector-borne disease may put a threat to the development of a society by absorbing the benefit of technological innovation (Caminade et al., 2014; Medlock and Leach, 2015). Thus, not only the already high infection rates of countries affected by the disease today are predicted to worsen in the future but countries which had never been interested by malaria will be affected by the latter and the associated vector-borne diseases in the next decades.

The rest of the paper is organized as follows. Section 2 presents the general setting and a review of the microeconomic and macroeconomic studies on the relationship between disease and economic development, discussing also about possible issues. In section 3 I provide a description of the data used in the analysis, while Section 4 presents the empirical strategy and results for farmer agricultural productivity. Section 5 examines the effects of the eradication of malaria on agricultural output per capita. Finally, Section 6 concludes.

2 Historical Background and Related Literature

In this section, I provide background information relevant to this study. First, I document the historical diffusion of malaria and associated vector borne diseases in the US along with climatic conditions which foster the transmission of the latter. This paper focuses on the effects of malaria eradication campaigns on US counties with different exogenous climatic conditions more or less suitable for the reproduction of mosquitoes larvae responsible for the diffusion of malaria. To this regard I describe the two indices of malaria prevalence used to perform the analysis (i.e. the Malaria Suitability Index and the Malaria Endemicity Index). Second, the paper analyses historical farm output and productivity levels for each US county. Moreover, since one of the purposes of this study is to evaluate the mechanisms through which the eradication of malaria impacts agricultural productivity, I also discuss the total amount of arable land at county level before and after the eradication of malaria. Causal evidence of the boost in agricultural output and the productivity of farmers due to better health conditions thanks to the newly available cure against malaria along with the associated increase in population is presented in Section 4 and 5, respectively.

2.1 Historical Malaria Prevalence and the Eradication Policies in the US

Malaria has always been an important factor of the economic development. As Figure 3 shows until early 1900s south western U.S. and Europe were largely affected by the disease. It was not until the 20^{th} century when technological advancements and the discovery of new drugs along with modern chemical components made possible to effectively prevent the transmission of malaria.

The first major discovery was quinine in the late 17^{th} century when it was imported first in Spain as a curative plant. Its beneficial effects were not fully understood for almost two centuries[4]. In 1897 Ronald Ross succeeded in demonstrating the life-cycle of the parasites of malaria in mosquitoes, thus establishing that malaria came from the bite of mosquitoes, for this discovery he was awarded the nobel prize in medicine. Consequently the properties of quinine were fully appreciated and the plant started being used in hospitals both in Europe and the USA as an effective drug in curing the fever caused by malaria (Majori, 2012). The first interventions to contrast the diffusion of malaria were adopted during early 1900s[5] which consisted mostly in the administration of quinine to people living in areas with high infection rates along with the drainage of swamps and wetlands. An example of the effectiveness of early control policies was the construction of the Panama Canal in 1910s. Yellow fever and

[4]Due to its bitter taste people were reluctant to consider it as a real measure against malaria

[5]Italy started an active fight against malaria in 1900 when the law ensuring the national production and sale of the antimalarial quinine, the so called "Chinino di Stato", was promulgated (Majori, 2012).

malaria were a major cause of death and illness among workers in the area. The Center for Disease Control Prevention (2010) shows that in 1906, there were over 26,000 employees working on the Canal. Of these, over 21,000 (i.e. more than 80%) were hospitalized for malaria at some time during their work. By 1912, there were over 50,000 employees, and the number of hospitalized workers had decreased to approximately 5,600 (i.e. 11%).
In the USA quinine was made available at every state hospital in the early 1900s thanks to donations of the Rockefeller Foundation and the newly established United States Public Health Services (USPHS) along with new methods such as the spraying of larvicides and wetland conversion. By 1912 all board of health of each state followed the USPHS model[6] (Williams, 1952; Humphreys, 2001; Saul, 2002). Thus, until early 1900s scientific advancements were not adequate to treat malaria. This caused more than 80% of the countries in the world to be affected by vector-borne diseases. Scientific progress which brought to the availability of modern drugs along with better sanitation and socio-economic conditions led to a shrink in the burden of malaria. In particular, the U.S. and southern Europe which were largely affected by vector borne diseases successfully eradicated the latter during the first half of the 20^{th} century. During the whole 20^{th} century to the present day the burden of malaria was constantly reduced. Bleakley (2010b) shows how during the twenty year period of early interventions against malaria (i.e. from 1900s to 1920s): mortality rates of malaria declined by more than 70%. However, this is only part of the story, since morbidity rates rather than mortality fully represents the true impact of malaria into the economy. After the first positive results of the fight against malaria until the 1920s, a small resurgence of the latter followed during the period ranging from the 1930s to the 1940s caused by the great depression and the second world war. Finally, the use of DDT and newly discovered chemical components caused a reduction of the burden of malaria to negligible levels during the early 1940s (Bleakley, 2003).
Digitizing historical maps Hay et al. (2004) have quantified the anthropogenic impact on the distribution of malaria in the 20^{th} century at six intervals between 1900 and 2002 (i.e. 1900, 1946, 1965, 1975, 1994, 2002) with 1900 being the last year before the first adoption of eradication policies.

As shown in Figure 3 the U.S. were also historically affected by vector borne diseases until the first half of the 20^{th} century. Before the implementation of vector control policies, the problem was so serious that, as shown in the 9^{th} Census of the United States realized in 1870, for some south western counties more than 10% of deaths from all causes were caused

[6] Williams (1952) presents a thorough history of the US Public Health Service. Humphreys (2001) summarizes the history of malaria-control efforts in the United States. The annual reports of the Rockefeller Foundation's International Health Board (1919) provide information about its anti-malaria demonstration projects. Much of the historical detail for the United States is drawn from these sources

by malaria alone. In other words 1 out of 10 people died because of causes linked to malaria. Figure 10 below shows in detail the proportion of deaths caused by malaria to the total number of deaths[7].

In the U.S. early effective measures to prevent the transmission of vector borne diseases started with the establishment of the U.S. Public Health Service (USPHS) and the programs implemented by the Rockefeller Foundation in early 1900s. Such measures lasted until 1930s when new technological advancements made the discovery of the DDT possible along with new fertilizers. The newly discovered chemical component, originally developed as an insecticide for use as an agricultural and household pesticide, further reduced the reproduction rates of mosquitoes and improved agricultural production Derryberry and Gartrell (1952). DDT was extensively used in agriculture from 1940s to early 1970s when it was finally banned in the US because of the controversial effects it had on the environmental biodiversity and human health. The US Environmental Protection Agency (EPA, 1975) has estimated that a total of 1.8 million tonnes have been produced globally since the 1940s with more than 600,000 tonnes produced in the U.S. only before the 1972 ban. The usage of DDT reached its peak in 1959 at about 36,000 tonnes in the United States only.
Given the aforementioned evidence malaria has possibly played an important role in the socio-economic development of different U.S. areas both in terms of population and labour productivity. Hence one might ask whether more endemic areas have suffered in terms of human capital compared to less endemic ones.

2.2 Historical Farmer Productivity in the US

Figure 9 shows the average annual farmer agricultural productivity[8] growth rates per US county for the period 1870 to 1900 on the left and for the period 1900 to 1920 on the right [9]. In Figure 4 below the respective correlation coefficients are displayed. The graph on the left indicates a marked negative correlation between farmer agricultural productivity growth prior to 1900 and malaria prevalence[10]. The correlation between farmer agricultural productivity growth and malaria prevalence appears to dramatically switch sign after 1900 as shown on the left hand side of Figure 4. In particular, correlation between farmer agricultural

[7]The map shown in Figure 10 has been created by Francis A. Walker using data from the Statistical Atlas from the 9^{th} Census of the United States 1870

[8]County farmer agricultural productivity is defined as the total farm output value per farmer for each county. More details on data and variables are provided in Section 3

[9]The map shown in Figure 9 was created using the geographic information system ArcGis software. Green areas have experienced a greater increase in farm productivity growth while areas in red have shown a decrease in farm productivity over time

[10]1870 is the first year for which farm productivity data per US county is available in Haines et al. (2005)

productivity growth between 1870 and 1900 and the index of malaria prevalence[11] was -0.36. This suggests that counties where malaria was more prevalent experienced lower farmer agricultural productivity growth prior to 1900. However, the sign of the correlation between farmer agricultural productivity growth and the prevalence of malaria appears to reverse to +0.17 after 1900 as shown in the right hand side of Figure 4. Specifically the average annual farmer agricultural productivity growth between 1900 and 1920 was higher in counties where malaria was historically more prevalent. Thus, suggesting that consequently to the active fight against malaria thanks to the discovery of new drugs and chemical components such as the wide use of quinine at each state hospital and spraying of larvicides (Williams, 1952; Humphreys, 2001) counties which where historically more affected by malaria experienced larger growth rates in farmer agricultural productivity compared to US areas less affected by malaria. Together, Figures 9 and 4 suggests that something happened in 1900 that has caused farmer agricultural productivity to grow more in areas historically more affected by malaria compared to areas in which malaria was less prevalent. Potentially the eradication of malaria, which happened between 1900s and 1920s, could possibly be a determinant of such increase in farmer agricultural productivity occurred in more "malarious" US counties.

2.3 Related Literature

To this regard, economic literature has extensively studied the relationship between disease with particular interest to malaria, and human capital loss for a long time. In particular, previous microeconomic literature has studied the impact of different disease control policies on human capital providing large evidence of the positive impact of disease prevention campaigns on *individual* education levels which are widely regarded as fundamentals of persistent economic growth Bleakley (2007), Bleakley (2010a), Cutler et al. (2010) and Bleakley et al. (2014). However, macroeconomic studies have generally emphasized the Malthusian view, arguing that the expansion in population in the short-term due to the increase in life expectancy could not be matched by an increment in the availability of natural resources, thus leading to modest (if any) improvements on GDP per capita (Hansen and Lønstrup, 2015; Acemoglu and Johnson, 2007; Ashraf et al., 2008).

No study, however has clearly addressed the role that malaria might indeed have played in explaining different historical economic development of US areas. This because some counties might have benefited from the eradication of malaria in terms of farmer agricultural productivity, therefore generating positive effects on their economies compared to less endemic counties.

[11] Detailed description of the indices used to measure the prevalence and transmission of malaria disease is provided in Section 3.1

Gallup and Sachs (2001) and Sachs and Malaney (2002) first pointed out the correlation between malaria transmission and economic growth arguing that reversed causality would not pose a substantial problem. However, from their first work a large strand of economic studies on the causal impact of health and disease on economic growth has followed, showing that people living in areas with higher levels of malaria infection prior to some eradication campaign experienced greater increases in school attendance and literacy afterwards compared to people living in non malarious areas in the Americas, India and Italy (Bleakley, 2010a; Cutler et al., 2010; Percoco, 2013). Specifically, Bleakley (2003) and Bleakley (2010b) show that level of income of adults not exposed to tropical diseases during their childhood was higher compared to those exposed to weather conditions suitable for transmission of infectious diseases.

However, as argued by Acemoglu and Johnson (2007) and later by Hansen and Lønstrup (2015) microeconomic studies are likely overestimating the real impact of disease transmission reduction since they do not control for general equilibrium effects of the increased life expectancy due to less mortality. By instrumenting life expectancy with the predicted mortality rates of 15 major diseases of the 20^{th} century Acemoglu and Johnson (2007) find a negative relationship between the latter and GDP per capita growth. Finally, in a recent macroeconomic study, by using highly disaggregated data Gooch (2017) shows the positive effect that malaria eradication has had on world population and population density. That said, macroeconomic studies only consider mortality rates rather than data morbidity rates. As a consequence, the economic burden of the infectious disease is likely to be underestimated since only a fraction of people affected by malaria effectively die, while a lot more find themselves physically debilitated thus not being able to work efficiently.

To have a comprehension of the possible effect of malaria on farmer agricultural productivity is important to notice that prior to any intervention against malaria malarious countries had large portions of their cultivable land actually uncultivated or barely cultivated because of the prevalence of mosquitoes in those regions. For instance, as reported by Brown (1986) in 1898 Fortunato and Franchaneti wrote a letter to their sponsor indicating the devastation caused by this disease in Italy. "*Malaria disease leaves uncultivated 2 million hectares of land*[12]*. It poisons every year about 2 million inhabitants and kills 15,000 of them. There is no other health problem so deeply linked to the prosperity of our country.*" This means that the eradication of malaria other than increasing life expectancy and in turn population[13] could have made the amount of land which could not be cultivated before (or only marginally cultivated) possible to cultivate without the prevalence of malaria, thus providing

[12]This corresponds to the equivalent area of the italian region Puglia

[13]Which might have a negative impact on the GDP per capita as shown in Acemoglu and Johnson (2007) and Hansen and Lønstrup (2015)

an exogenous positive output shock which might have counterbalanced the negative effect of the increased population.
By using a robust Difference in Difference approach (DID), this study aims at identifying the effects of the successful eradication of malaria in the U.S. on farmer agricultural productivity for each county. In other words since malaria historically strikes rural areas more than urban ones and mostly farmers, its eradication could have fostered farm activities and crop cultivation in those areas which were suitable for agriculture but highly endemic. Thus, a positive impact of malaria eradication on farmer agricultural productivity would possibly counterbalance the increase in population due to the lower mortality rate, meaning that we could “escape” from the Malthusian trap as pointed out by Acemoglu and Johnson (2007).

None of the above mentioned related studies has yet attempted to evaluate the impact of the eradication of malaria on the historical farmer agricultural productivity of each county in the US. As presented in Section 1 establishing a clear direction in the relationship between malaria and agricultural productivity is crucial to understand the causes of historical economic development of different areas of the US. Moreover, it is also important to evaluate if the eventual impact of malaria control on historical agricultural productivity was due to a mere increase in arable land (i.e. due to wetland conversion to cropland) or if the latter was explained by an increase in labour productivity deriving from better health conditions of farmers whom were not affected by malaria any longer. To this regard the present study investigates the two principal channels through which malaria control might effect agricultural productivity. These are:

1. Higher labour productivity due to better health conditions of farmers;

2. Greater availability of cropland due to wetland conversion to arable land.

Therefore, this study aims at answering two questions: has the eradication of malaria increased farmer productivity of endemic counties with respect to less endemic ones? And, how much of the eventual increase was due to the mere increment of arable land and to the higher labour productivity of farmers? The robust empirical strategy in this study allows to disentangle the two effects of malaria control policies on farmer productivity.

3 Data

To perform the empirical investigation I construct the treatment variable by using two time invariant indices measuring the stability and the endemicity of malaria, respectively. The first index considers climate variables only which make the survival of mosquitoes larvae more or less favorable. The second index instead, considers the actual prevalence of malaria prior to the beginning of the eradication policies in the US (i.e. before early 1900). I interact the post malaria eradication policies indicator with the two measures of whether counties are suitable for the transmission of malaria. The dependent variable is constructed using historical disaggregated data on US farm output at county level. This section briefly explains how each variable is constructed or obtained. Time series gridded data on total amount of historical arable land are also used to assess whether the eradication of malaria increased the available cropland. Finally, data on county population allows me to explore whether the eradication of malaria beside having an effect on the agricultural output and productivity also affected population of more malarious counties. This data is essential to understand if the eventual benefits on agriculture due to the cure of malaria were offset by the increase of total population.

3.1 Measuring Spatial Prevalence of Malaria

First, highly malaria endemic counties were identified by using the Malaria Endemicity Index (henceforth, MSI). The MSI was first developed by Kiszewski et al. (2004) and is a spatially disaggregated time-invariant global index representing the stability of malaria transmission. Biologically, the MSI measures the suitability of particular climatic conditions which foster the reproduction of two particular species of mosquitoes which are the natural vehicle of the most severe category of malaria[14] Figure A**??** in the appendix, shows how regions such as Sub-Saharan Africa, Latin America and South East Asia present the ideal climatic conditions for the reproduction of species of mosquitoes transmitting malaria to human beings. However, historically malaria was also prevalent in many other regions of the world. Notable examples of regions in which malaria was widespread are: the Western US, Southern Europe and Northern Australia. Figure A**??** is at 0.5°latitude x 0.5°longitude gridded level which corresponds approximately to 56km x 56km at the equator. To get the average malaria stability value for each US county spatial data on malaria is then intersected with historical US county borders[15]. This procedure will generate a time invariant malaria stability index for each county based upon climatic conditions only[16].

The MSI index has been recently used in a number of recent papers in development economics (Alsan, 2015; Giuliano and Nunn, 2013; Michalopoulos and Papaioannou, 2013; Easterly and Levine, 2016; Henderson et al., 2017) for its advantage of being independent

[14] The asymptomatic, uncomplicated categories of malaria are not considered in this study.

[15] Spatial data on historical US county borders are retrieved from the Newberry of historical county borders of the Newberry Library. The historical US county borders contains publicly available spatial data on counties from 1629 to 2000

[16] Since some counties had an area smaller than 0.5°latitude x 0.5°longitude I first, use the natural neighbor algorithm to interpolate the malaria stability values. The interpolation method allows to lower the scale to 0.1°x 0.1°corresponding to an area approximately equal to 5 km x 5 km at the equator.

of socio-economic conditions. Using a time-invariant index representing the suitability of each county to the transmission of malaria and related vector borne diseases is crucial to alleviate potential endogeneity issues. Indeed, one of the benefits of the MSI is that it does not directly measures human infection rates which might in turn depend upon socio-economic conditions. The MSI is indeed constructed by using weather variables available at 0.5°latitude x 0.5°longitude level[17] along with the estimation of the survival time of mosquitoes larvae based on regionally dominant species of Anopheles mosquitoes. Another benefit of the MSI is that it is precise, since it is based on information measured with contemporary GIS accuracy. This may explain why the stability index captures moderate malaria prevalence more effectively.

A second measure of historical prevalence of malaria is the index developed by Hay et al. (2004) by digitizing historical maps shown first in Lysenko and Semashko (1968). However, this second malaria stability index is less precise that the MSI, since it is constructed at 1°by 1°, and most importantly it uses actual infection rates data, thus potentially being influenced by human activity. For these reasons, this study uses both the MSI and the MEI as indexes of *suitability* of transmission of malaria and actual *prevalence* of the disease.
The MSI is a continuous variable whose values range from 0 to 39 with 0 being totally malaria free grid (i.e. the mosquito larva cannot survive with these climatic conditions) and 39 the grid with weather conditions most suitable for a stable and permanent transmission of malaria (i.e. mosquitoes larvae may survive and reproduce without obstacles of climatic nature any day of the year). However, the stability index comes as a categorical version too[18]. Therefore, in the empirical analysis I use the MSI index as both a continuous and discrete variable.

The MEI is also a continuous variable capturing the observed distribution of malaria. MEI values range from 0 to 1 with 0 being a grid for which no case of malaria was observed until 1900 and 1 being a grid with the highest infection rate of malaria. The lowest endemicity level is hypoendemic with a value ≤ 0.1, followed by mesoendemic grid ≤ 0.5, hyperendemic with a value ranging from 0.5 to 0.75 and holoendemic with a value > 0.75. Although the MEI index ranges from 0 to 1 worldwide, values for the USA do not exceed 0.5 (in other words the USA ranged from being totally malaria free to mesoendemic, while cases of hyperendemic and holoendemic areas were not registered).

[17]Weather data used for constructing the MSI are monthly average, minimum and maximum temperature, humidity and precipitations. Climatic data are exclusively provided by the Climate Research Unit at the University of East Anglia

[18]Grids with values from 0 to 0.05 are considered as malaria free while grids with values greater than 0.06 present climatic conditions suitable for a certain persistence of malaria transmission. Kiszewski et al. (2004) differentiated stability levels in 9 categories: 0-0.05, 0.06-1, 1.01-2, 2.01-5, 5.01-8, 8.01-12, 12.01-18, 18.01-25, 25.01-39

Since this study merges spatial data on historical county boundaries with 0.5°latitude x 0.5°longitude gridded data on malaria stability and endemicity values, some historical counties had an area which was less than that of a grid. To fix this issue the MSI and the MEI are interpolated at a very disaggregated geographic level (i.e. 0.2°latitude x 0.2°longitude) in order to have at least an stability value for each county[19]. I then calculate the mean of the index for each county. The results from the intersection of the MSI and the MEI and the historical US counties are shown in Figures 1 and 2. Important to notice is that the Malaria Stability Index (MSI) relative to the US territories ranges from 0 to about 1.5. This means that the transmission of malaria was largely unstable, implying that climatic conditions suitable for the survival and reproduction of mosquitoes larvae were mainly concentrated in one season, e.g. summer[20].

3.2 Farmer Productivity and Total County Population

Since malaria disease mostly affects rural areas and farmers, we expect that a direct consequence of the eradication of vector-borne diseases is the improvement in farmers health. I use farmer agricultural production as a proxy for health of farmers. Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 (Haines et al., 2005) provides for two different farmer productivity measures for each US county, namely the total farm value and farm output per farmer.

Figures 13 and 14 in Section 3.1 provide a visual representation of the correlation between the two measures of malaria incidence and farm productivity throughout the 20^{th} century. Farm value per county farmer is defined as the value of all farmland, housing and outbuildings at the time of census enumeration divided by the number of farmers in a specific county. As an alternative measure of farmer productivity, I use farm output per farmer, defined as the total value of all farm products, such as crop and livestock products, within the year prior to the enumeration day divided by the number of farmers. Unfortunately, farm output is not available in the census years of 1910, 1920, and 1930. Therefore, I consider the period 1870 to 1900 as pre-eradication of malaria while 1940 to 1970 as post eradication period. Historical, Demographic, Economic, and Social Data: The United States, 1790–2002 provides data on farm value per county registered at each agricultural census starting from 1850 and undertaken every decade until 2000. However, the quality of agricultural census was acceptable only starting from 1870. This determined the choice to use farm value and farm output per farmland county acre and farm value and farm output per farmer only starting

[19] I have used the natural neighbor interpolation method automatically computed by the GIS software.

[20] Higher MSI values do not necessary imply a large number of people affected but more stable climatic conditions for species of anopheles mosquitoes to live throughout the whole year.

from 1870. Therefore, pre-eradication period starts from 1870 until 1900 with 1910 being the starting year of vector control policies which took place until 1940s with the introduction of the DDT. A worthwhile consideration is the change in county boundaries during the period 1870-1970. These changes of county boundaries matter when county fixed effects are considered. As in Bleakley and Hong (2017) to partially fix the problem, I adjust farm value, along with all the other variables at county level that I control for, on the 1870 county boundary using the area-weighted average method. Figures 13 and 14 shows respectively the MSI index and the MEI joined with the log farm value per farmer in different decades before and after the introduction of vector control policies. Panel A, B and C show a significant negative correlation between the MSI and the MEI and the log of farm value per farmer which does not appear to change until 1900 (i.e. prior to start of the eradication campaigns) with associated slopes being -0.25, -0.28 and -0.27 for 1880, 1890 and 1900 respectively (-0.14, -0.19 and -0.22 for the MEI). This negative correlation decreased in 1920 as shown in Panel D and disappears from 1950 until the present day as shown in Panels E and F with associated slopes being -0.18, -0.12 and -0.07 for 1920, 1950 and 1970 respectively (-0.13, -0.09 and 0.09 for the MEI).

I also use time series data on total county population from Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 (Haines et al., 2005).

3.3 Arable Land

Finally, I examine the effects that the eradication campaigns against malaria in the US had on the total amount of arable land using time series gridded data from Klein Goldewijk et al. (2017). I compute the proportion of cropland out of total county-acre starting from 1870 for each decade until 2000. Figure 3.3 shows that the conversion from wetland to arable land was not significant in more malarious counties after the eradication of the vector-borne diseases.

4 The Direct Effect of Malaria Eradication on Farmer Productivity

In this section, I examine the causal effects of the eradication of malaria in the US on farmer productivity. First, I describe the econometric strategy in which the treatment is defined as the interaction between the exogenous county variation in weather conditions more or less suitable for the transmission of malaria disease and a post eradication indicator variable. Using this identification strategy I am able to assess the relationship between malaria and agricultural productivity of farmer along with the associated mechanisms (i.e. better farmers' health and greater availability of arable land due to land conversion from swamps and wetland into cropland).

4.1 Empirical Framework

The estimation method presented below follows the same logic of a standard differences-in-differences (DID) strategy. This allows for the comparison of counties with high levels of stability of malaria with counties with low malaria stability and between counties with high historical incidence of malaria and counties with low incidence of the vector borne disease. The difference between the empirical strategy presented in this study and the standard DID is that the treatment (which in this specific case is represented by the interaction between a post treatment indicator variable and one of the two indices measuring the incidence of malaria) is not discrete but it is a continuous measure (namely, I do not compare malarious counties with non malarious ones but more malarious with less malarious counties). Therefore, more variation in the treatment is captured. The regression presented below estimates the impact of the county average MSI and MEI on the proportion of farmers per county-acre.

$$y_{i,s,t} = \alpha \; Malaria_i \cdot I_t^{Post} + \sum_{t=1870}^{2000} \Omega \mathbf{X}'_{i,s} I_t + \sum_{t=1870}^{2000} \delta_t I_t + \sum_p \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_q \omega_q I_t^q + \epsilon_{i,s,t} \tag{1}$$

where index i represents each US county in state s, and t indexes time periods considered in the analysis (i.e. 1870, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990 and 2000). Years from 1870 to 1900 are part of the pre-eradication period, and since first national-wide policies were implemented soon after 1900, 1900 is considered the last year prior to the interventions. Post intervention period ranges from 1910 until the 1950s when the last round of the National Malaria Eradication Programm (1951) was undertaken (US Center for Disease Control and Prevention, 2010). By the end of 1950s and early 1960s malaria was considered eliminated in all US counties also thanks to the usage of newly discovered chemical components such as the DDT[21]. The dependent variable is represented as $y_{i,s,t}$ which is the natural log of the two measure of farmers productivity, i.e. farm output per farmer and farm value per farmer per county i, state s at time t. The variable $Malaria_i$ is one of the two indices representing the incidence of malaria, i.e. the average time invariant malaria

[21]In 1945, the Dichlorodiphenyltrichloroethane, commonly known as DDT, was made available to farmers as an agricultural insecticide and played a role in the temporary elimination of malaria in Europe and North America. In 1972 the US lifted a ban on its usage due to its adverse effects on the surrounding environment as well as on human health.

stability index of county i and the average malaria endemicity index of county i. Variable I_t^{Post} is a post treatment dummy variable which takes value 1 for years after 1900 (i.e. 1910, 1920, 1930 and so on until 2000) while value 0 for years before the exogenous interventions against malaria (i.e. 1870, 1880, 1890 and 1900). This specification also includes county fixed effects $\sum_p \gamma_p I_t^p$, where p indicates the set of US counties; time period fixed effects $\sum_{t=1870}^{2000} \delta_t I_t$. $\mathbf{X}_{i,s}$ represents vectors of time invariant county-specific agricultural controls included in the regression. As county-level controls I use a set of relevant geographical and historical county specific characteristics which might have affected farm output and farmer productivity during the 20^{th} century as in Bleakley and Hong (2017). These are the following: county population density in 1870, the ratio of white population to county population in 1870, the ratio of farmland to total available county area in 1870, the proportion of farmers reporting the use of fertilizers in 1870, the proportion of farmland that reported have received drainage in 1870, the proportion of farmland with improved land in 1870 and county average agriculture suitability. A correlation matrix between the MSI and the MEI index and these population and agricultural controls is shown in Table A9 in the Appendix. Table A9 also shows the correlation between the MEI index and the MSI, the population density and the proportion of farmers out of total population. To account for these differences, when comparing malaria-free counties with more endemic counties, I include the following covariates: (1) two interaction terms: ln(Population Density in 1870) interacted with the MSI and ln(Proportion of white farmers out of total farmers in 1870) interacted with the MSI, and the stability index interacted with the full set of time fixed effects. Finally, in order to address the concern that the econometric strategy might simply capture the fact that southern states of the US might on average be more suitable for agriculture than northern states of the US and that each state might have had agricultural productivity growth rates differently from others because of any other reason besides the eradication of malaria I include state fixed effects interacted with time period fixed effects to the baseline specification presented in equation 1. State-by-time fixed effects are represented as follows: $\sum_{t=1870}^{2000} \sum_q \omega_q I_t^q$. Where q indicates the set of US states. With the state-year fixed effects, our coefficient of interest α is identified from within-state variation only, that is the empirical strategy compares counties within the same state and therefore subjected to the same federal policies[22].

Although I am aware of the fact that other variables not included in the regressions could have explained the historical changes in the relationship between malaria stability and farm productivity, the main limitation to the use of a higher number of controls is

[22] To this regard it is important to stress that decisions to tackle malaria along with other vector borne diseases (i.e. supply of free quinine, availability of newly discovered pesticides etc..) were made at the governmental level and implemented at federal level by each state. Counties had not power in deciding when and where to implement malaria eradication campaigns.

due to the long time period of the analysis. That said however, the inclusion of other control variables such as government policies and investments which might have had a role in explaining changes in farm productivity would absorb part of the effects of malaria treatment to agricultural productivity because government policies, for instance, are investments in adapting to circumstances, including the effects of malaria and other vector borne diseases. For this reason we would not want to remove such adaptation from the estimates, because they are part of the story. I use the same set of control variables at the county level as in Bleakley and Hong (2017) including also the average county suitability for crop cultivation. Agricultural Suitability Index (henceforth ASI) is computed by the Center for Sustainability and Global Environment at the University of Wisconsin-Madison (Ramankutty et al., 2002). To compute the suitability for agriculture Ramankutty et al. (2002) rely on weather and environmental conditions required for crop cultivation. Weather conditions are taken from the global climatic database compiled by the Climate Research Unit at the University of East Anglia[23].

4.2 Addressing Potential Threats to Validity

A potential threat to the validity of the empirical analysis comes from the possible endogeneity of the treatments to malaria which could be adopted by highly malarious counties first. In other words counties with greater diffusion of malaria could have received the treatment before counties with low or no infection rates. If this is the case, then having received the treatment before less malarious counties would bias the results upwards. As discussed in Section 1, economic literature addressing the effect of diseases and health on economic outcomes (Bleakley, 2007; Cutler et al., 2010; Gooch, 2017) and more recently on conflicts (Acemoglu et al., 2020) has argued that the reduction in the burden of malaria clearly resulted from critical scientific innovations coming overwhelmingly from outside the highly endemic counties and which have culminated with better understanding of the origins of the disease and the discover of new drugs and chemical components. Therefore, major discoveries on malaria such as that it is transmitted to human beings through the bites of certain species of mosquitoes living in particular weather conditions or the finding of the curative properties of quinine and the diffusion of newly discovered pesticides and chemical components did not come from regions in which malaria was more prevalent. Moreover, as shown in Williams (1952) and Humphreys (2001) decisions on the adoption of campaigns addressed to the eradication of malaria were taken at the governmental level and implemented at the federal level by each state. Based on historical evidence we conclude that the eradication of malaria depended

[23]The database includes monthly data on weather conditions (e.g. precipitations, cloud cover, minimum, maximum temperature, humidity etc.) from 1900 until 2016.

on discoveries from abroad and counties had no decision power in terms of availability and administration of quinine to people affected from malaria, use of pesticides and later on of the DDT and drainage of swamps and wetland. Hence, this attenuates possible concerns of endogeneity of the treatment by areas of the US more affected by vector borne disease. That said, I adopt a series of tests to check that:

1. The relationship between the two measures of prevalence of malaria and agricultural productivity per county did not change before 1900;
2. Counties with high incidence of malaria are comparable to counties with low incidence of the disease;
3. The cutoff date (i.e. 1900) is reasonably correct and no other cutoff produces statistically significant results.

4.2.1 Flexible Estimates

As with a standard difference-in-difference, the empirical strategy adopted in this study relies on the assumption that no other event, besides the availability of new drugs and effective chemical components in preventing the transmission of malaria and vector borne diseases, also occurred in early 1900s and affected farm productivity differently from each US county. This is a crucial assumption which should not be taken for granted since the US have experienced many changes during the 20^{th} century. This issue is partially fixed by the massive number of counties considered in this analysis[24]. However, I implement a number of cautions in order to examine whether the patterns in the data are consistent with this assumption. First, as described in section 2 the historical evidence suggests that in the US early effective measures act to prevent the transmission of vector borne diseases started with the establishment of the US Public Health Service (USPHS) in early 1900s. Those measures consisted in the free availability of quinine in all hospitals in the US, the use of newly discovered pesticides along with the drainage of swamps and wetland. Given this evidence, the most reasonable cutoff date is 1900, and therefore 1910 is the first post-adoption time period. Second, in order to exclude that no event other than the eradication of malaria happened from 1900s and affected farm productivity differently for each US county, I estimate a fully flexible estimating equation as in Nunn and Qian (2011), which takes the following form:

[24]almost 2000 historical counties are part of the analysis

$$y_{i,s,t} = \sum_{j=1870}^{2000} \alpha_j \; Malaria_i \cdot I_t^j + \sum_{t=1870}^{2000} \Omega \mathbf{X'}_{i,s} I_t + \sum_{t=1870}^{2000} \delta_t I_t + \sum_p \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_q \omega_q I_t^q + \epsilon_{i,s,t} \tag{2}$$

The only difference from equation 1 is that in equation 2, rather than interacting $Malaria_i$ with a post-adoption indicator variable, I interact the MSI and the MEI with each of the time-period fixed effects. The estimated vectors of α_j now reveal the relationship between the MSI (and the MEI) and farmer productivity per county-acre in each time-period (e.g. we will have an estimate α_j for each decade from 1870 to 2000). If, for instance, no other event occurred in early 1900s and the eradication of malaria had a positive effect on farm productivity, then we would expect the estimated α_js not to be statistically significant over time for the years before the adoption of vector control policies while becoming positive and statistically significant after the eradication campaigns started in early 1900s. We would also expect the α_js to be constant after the eradication campaigns successfully ended (i.e. 1960s circa)

Figures 5 and 6 show the estimated α_js compared to 1870 (i.e. the starting year of the analysis) along with the associate confidence intervals.

4.2.2 Neighboring Counties Analysis

Furthermore, I limit the analysis by only comparing *neighboring* counties belonging to the same state (and therefore subject to the same federal policies) which had substantially different MSI and MEI values. In particular, when considering the MSI as variable representing the diffusion of malaria, I compare each county showing weather conditions suitable for the transmission of malaria (i.e. with a MSI ≥ 0.06) with its neighboring counties[25] that presented weather conditions not suitable for the reproduction of mosquitoes larvae (i.e. with a MSI ≈ 0). This restriction substantially reduces concerns of any difference between counties with more or less malaria stability. Therefore, the procedure allows to select only those neighborhoods containing at least one malarious and one malaria free county [26]. The reasoning behind this restriction is that we would be sure to compare counties which are very similar, that is they are close enough, and belong to the same state but only differ in terms of malaria stability values.

[25](Neighboring counties are defined as those counties having their centroids within a grid of 2° latitude by 2° longitude and belong to the same state)

[26]For instance a neighborhood containing all malarious counties was not considered. The same reasoning was applied to neighborhoods with no malarious county. In this way we do not compare a malarious county in Alabama with a malaria free county in Montana

For the same reason, when considering the MEI as variable representing the diffusion of malaria, I compare each county showing high endemicity of malaria (i.e. with a MEI ≥ 0.1, or in other words mesoendemic counties) with its neighboring counties that presented showed no diffusion of malaria (i.e. with a MEI ≈ 0 or malaria-free counties).

A total of 258 counties were obtained following the aforementioned restriction[27]. I therefore, create an indicator variable I_n for each neighborhood n considered which is added in equation 1 so to compare endemic counties with non-endemic counties belonging to the same neighborhood. The estimated regression therefore becomes the following:

$$y_{i,s,t} = \alpha \; Malaria_i \cdot I_n \cdot I_t^{Post} + \sum_{t=1870}^{2000} \Omega \mathbf{X'}_{i,s} I_t + \sum_{t=1870}^{2000} \delta_t I_t + \sum_p \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_q \omega_q I_t^q + \epsilon_{i,s,t} \tag{3}$$

Where $Malaria_i$ is again one of the two indexes, namely MSI based on weather conditions suitable for a stable reproduction of mosquitoes larvae and the MEI based on actual infection rates of malaria.

In equation 3 the coefficient of interest α is identified no more from within-state variations, but *within-neighborhood* variation only. With this restriction the econometric procedure not only compares counties which are belonging to the same state (and therefore subjected to the same federal policies) but also neighboring counties. If highly malarious counties were effectively different from less malarious counties we should expect the parameter estimates of years prior to the implementation of vector control policies to be statistically significant. However, as Figure 7 and 8 show this is not the case. Results obtained when considering the almost 2000 counties are thus confirmed by the restriction of neighboring counties.

4.2.3 Placebo Treatment Periods

The estimated coefficient for the interaction between the Malaria Endemicity Index and the post eradication dummy variable indicates the average increment in farm productivity between the pre and post periods for highly malarious counties relative to less malarious counties. As shown early, it is crucial that in order to be valid, estimates, are expected to be close to zero until the cutoff period. I have chosen 1900 as cutoff year since it is consistent with the historical description of the discovery that malaria was transmitted to humans through the bite of mosquitoes and of the curative properties of quinine with its subsequent mass distribution. Therefore, until 1900 there is no reason to expect counties

[27]The 258 selected counties belonged to a total of 119 different neighborhoods

with higher prevalence of malaria to have differential growth in farm productivity.

4.3 Results

Table 1 reports different specifications of equation 1 where the variable of interest is the suitability of malaria, i.e. the MSI. Therefore, the results reported in Table 1 show the effects that the eradication of malaria after 1900 had on the farm value per farmer of counties with initial high weather suitability for the transmission of malaria compared to counties with initial less suitability for malaria. Column (1) includes only state-specific linear trends, time and county fixed effects. A a unit increase in the stability of malaria index is associated to an increase of almost 20 percentage points in farm productivity, measured as farm value per farmer, of counties with weather conditions more suitable to the transmission of malaria. Moving across columns more controls are added, including controls on population density which might have pushed for a greater agricultural productivity, the proportion of white people out of total county population, the use of fertilizers, drainage of land and the use of improved land along with the average agricultural suitability of county i. Nevertheless the coefficients indicating the impact of the MSI on farm productivity do not appear to considerably change in magnitude and of statistical significance thus, reducing concern for selection on unobservables (Altonji et al., 2005). The preferred specification is reported in column (4) of Table 1 and includes all population and agricultural controls other than including state-specific linear trends, time and county fixed effects. Results in column (4) show that after the implementation of malaria eradication policies a a unit increase in the MSI has increased farm productivity of highly malaria suitable counties by 18 percentage points with respect to less malaria suitable counties. Table 2 instead reports the results of the impact of the MSI on the second measure of farm productivity, i.e. farm output per farmer. Here, the results observed in Table 1 are confirmed. Therefore, the positive impact of the stability of malaria after its eradication is consistent with both measures of county agricultural productivity i.e. farm value per farmer and farm output per farmer. Similarly Table 3 and Table 4 report the results of equation 1 where variable $Malaria_i$ is the average endemicity value (MEI) for each county i. Table 3 shows the impact of the MEI on county farm value per farmer, while in Table 4 results of the impact of the MEI on county farm output per farmer are reported. The preferred specification is reported in column (4) of both Table 3 and Table 4. Estimates, are consistent with those in Table 1 and Table 2. We, thus confirm that the eradication of malaria, which in the US took place in early 1900s, caused an increase in agricultural productivity in counties were malaria was more prevalent and stable.

4.3.1 Results with Flexible Estimates

Specifically, Figure 5 shows the relationship between the time invariant MSI index (on the left) and the MEI index (on the right) and the first measure of agricultural productivity per farmer, namely the natural log of farm value per farmer. Figure 6 instead uses a second measure of agricultural productivity per farmer, namely the natural log of farm output per farmer.

Taken together, Figures 5 and 6 show a clear pattern, i.e. the relationship between malaria incidence and farm productivity for each US county is not statistically significant different from zero until 1900 (i.e. last year prior to the onset of the eradication policies). A spike between 1910 and 1920 emerges, followed by a small resurgence in during the 1930s and 1940s and then it steadily increases in magnitude from 1940 to 1970. Coefficients shown in Figures 5 and 6 is also crucial to exclude that any other event except of the eradication of malaria occurred during the time periods immediately prior to the implementation of vector control policies. These results show that counties with a greater degree of incidence of malaria were no different in terms of agricultural output from counties with less prevalence of malaria prior to the eradication of the vector borne disease and that after the eradication of malaria more endemic counties showed an increase in farmer productivity relatively to less endemic ones.

As Figures 5 and 6 show, after 1900 farmer productivity of counties with higher incidence of malaria begins to increase relatively to counties which were less endemic and less suitable for the transmission of vector borne diseases.

A potential limitation of this procedure is that from Figures 5 and 6 the mechanism underlying the increase in farmer productivity of more endemic counties cannot be clearly established. In other words: how much of the latter was caused by the increase in cropland, in turn due to land conversion, and by the greater labour productivity of farmers, in turn due to better health conditions? This question will be addresses in the next section.

4.3.2 Results of Neighboring Counties Analysis

Specifically, Figures 7 and 8 shows respectively the estimates of the impact of the MSI and the MEI on the measures of agricultural productivity relative to a baseline time-period, which we take to be 1870. Therefore, the absolute level simply tells us the difference in the relationship relative to an arbitrarily chosen baseline. We then would expect the parameter estimates of 1880, 1890 and 1900 (i.e. years prior to the discovery of anti malaria methods) not to be statistically significant different from the baseline year, while being positive from 1910 onward. What we observe is that prior to the eradication of malaria the relationship between the diffusion of malaria and agricultural productivity was not statistically different from highly malarious counties compared to less malarious counties. After 1900 however, this relationship starts to be positive until 1920s, indeed during 1930s and 1940s we observe a resurgence of the transmission of malaria followed by a stable decrease in the number of infected people which translates into higher growth rates of agricultural productivity.

Estimates of equation 3 are reported in Table 5 and Table 6. Specifically, Table 5 reports the results of the estimation of equation 3 in which the dependent variable is represented by the natural logarithm of county average farm value per farmer, while Table 6 reports the results of the estimation of equation 3 in which the dependent variable is represented by the natural logarithm of county average farm output per farmer. Column (1) includes only state-specific linear trends, time and county fixed effects. Results here are very similar in magnitude compared to the estimates of equation 1 and show that counties with weather conditions more favorable to a stable transmission of malaria (i.e. counties whose MSI ≥ 0.06) experienced an increase of more than six percentage points in farm productivity compared to malaria free neighboring counties (i.e. counties whose MSI ≈ 0). As before, concerns for selection on unobservables are also mitigated by the fact that moving across columns coefficients indicating the impact of the MSI on farm productivity do not appear to considerably change in magnitude and of statistical significance. The preferred specification is reported in column (4) and includes controls on county population density in 1870, the ratio of white population to the total county population in 1870, the ratio of farmland to total available county area 1870 the proportion of farmers reporting the use of fertilizers in 1870, the proportion of farmers that had their farm drained in 1870, the proportion of farms with improved land and the county average agriculture suitability index, other than including state-specific linear trends, time and county fixed effects. Results in column (4) show that counties with MSI ≥ 0.06 experienced an increase of about seven percentage points in farm productivity compared to neighboring counties with MSI ≈ 0 after the eradication of malaria.

Similar tables are produced showing the results of the estimation of equation 3 considering the MEI and are included in the Appendix. Here, we compare counties with high actual infection rates (i.e. counties with MEI ≥ 0.1) to again neighbor counties which had no cases of malaria (i.e with MEI ≈ 0). Results obtained considering the actual infection rates of malaria are consistent with those obtained with the time invariant malaria stability index.

4.3.3 Results of Placebo Treatment Periods

To test this assumption I estimate equation 1 considering different placebo cutoff dates. The estimates are reported in Table 7. Column (1) reports estimated effects for farm productivity using a sample that includes forty-year pre-eradication periods, ranging from 1870 to 1900 (i.e. 1870, 1880, 1890 and 1900). For this regression, the post indicator variable I^{Post} takes on the value of zero in 1870 and the value of one in 1880, 1890 and 1900. Column (2) reports estimated effects for farm productivity which consider the same period of column (1) but with the post indicator variable I^{Post} taking on the value of zero in 1870 and 1880 and the value of one in 1890 and 1900. Therefore, in columns (1) and (2) I check the validity of 1880 and 1890 respectively as cutoffs. Since the major discoveries of the treatment against malaria and in turn, the first adoption of vector control policies in the US did not take place before 1900 the results from specifications in columns (1) and (2) can be interpreted as a placebo experiment. To this regard, results in columns (1) and (2) yield statistically not significant coefficient estimates confirming that endemic counties did not have differential growth in farm productivity throughout the period from 1870 to 1900. Moreover, columns (1) and (2) prove that 1890 and 1900 cannot be considered as plausible cutoffs. Column (3) instead reports estimates using a sample that spans across forty years, ranging from 1890 to 1920. For this regression, the post indicator variable I^{Post} takes on the value of zero in 1890 and the value of one in 1900, 1910 and 1920. Here, all three of the post decades coincide with the postadoption period. The post indicator variable now coincides exactly with the postadoption period. This time, the results displayed in column (3) yield positive and statistically significant coefficient estimates. Similarly column (4) reports estimates using the same sample of column (3), ranging from 1890 to 1920. For this regression, the post indicator variable I^{Post} takes on the value of zero in 1890 and 1900 while the value of one in 1910 and 1920. As in column (3) results reported in column (4) yield positive and statistically significant coefficient estimates. Finally columns (5) and (6) report estimates using a sample that spans across forty years, ranging from 1940 to 1970. In column (5), the post indicator variable I^{Post} takes on the value of zero in 1940 while the value of one in 1950, 1960 and 1970 while in column (6) the post indicator variable I^{Post} takes on the value of zero in 1940 and 1950 and the value of one in 1960 and 1970. None of the two specifications reported in columns (5) and (6) yield to statistically significant results. Thus, confirming that the differential growth in agricultural productivity between highly malarious counties with less malarious counties happened after 1900 until 1940, and neither before or after that period.

The same exercise is shown in Table 8. Here, I consider the MEI as a measure of the transmission of malaria for each county. Again, as in Table 7 columns (1) and (2) do not yield statistical significant results. This is crucial to clearly state that the relationship between the endemicity level and the agricultural productivity of US counties did not change prior to 1900. Columns (3) and (4) show instead a slight negative estimate and a positive estimate respectively. This suggests that again more malarious counties experienced higher agricultural productivity growth rates only after 1900 and not before also when considering the actual data on infection rates of malaria. Finally, columns (5) and (6) of Table 8 also yield positive estimates. Therefore, it appears that the positive effect of the eradication of malaria on agricultural productivity in counties with higher infection rates lasted also after until 1970s.

Tables 9 and 10 show the same robustness check performed in Tables 7 and 8, this time considering the natural logarithm of the average county farm output per farmer as dependent variable. Farm output per farmer data at county level are missing for the years 1910, 1920 and 1930 therefore in Tables 9 and 10 I only check the non existence of any other event prior to 1900 which changed agricultural productivity of more malarious counties compared to less malarious one. Results shown in Tables 9 and 10 confirm the conclusions drawn in Tables 7 and 8.

Taken together, the results shown in Tables 7 and 8 confirm what shown in Figures 5 and 7, namely that the relationship between the prevalence of malaria measured both as stability and endemicity and agricultural productivity starts to changes only after 1900 and in no other previous year (i.e. 1870-1900).

4.4 Other Potential Mechanisms behind the Eradication of Malaria and Agricultural Productivity

Was the increase in farm productivity the effect of the increment in the availability of cropland due to the conversion of wetland? In this section, I provide evidence that the increase in agricultural output and farm productivity are not the result of the drainage of swamps and conversion from wetland to arable land. There is evidence that one of the principal measures against the diffusion of malaria was the drainage of swamps and wetland and the relative conversion of the latter into cropland (Snowden, 2008; Majori, 2012). However, if the agricultural output results were merely driven by the increase in cropland in counties with higher initial incidence of malaria we should find a significant relationship between malarious conditions and the total amount of cropland for each US county. In addition to that, there is no reason to believe that counties with more favorable exogenous malarious conditions converted greater amount of wetland into cropland after 1900. Nevertheless, using time series gridded data on cropland from 1870 until 2000, I examine whether malaria eradication campaigns increased the portion of arable land in counties with initial higher incidence of malaria.

Figure 11 shows no correlation between the diffusion of malaria and the available cropland per county neither before the eradication campaigns nor after their implementation. This leads us to have a clear insight on the main channels through which the control of malaria has increased farm productivity since it appears that highly endemic counties did not see a significant increase in their total cropland. In other words the increase in agricultural productivity in malarious counties was not the result of a greater availability of arable land which did not increase with the eradication of malaria but the result of a minor number of farmers needed in one acre of farmland. This means that if until 1900 in more malarious counties more farmers were needed in order to compensate for those who were affected by malaria, after its eradication the farm output per farmer increased compared to less malarious counties.

5 The Effects of Malaria Eradication on Farm Output per capita

I next turn to an investigation of the effects of the eradication of malaria in the US on natural resources per capita. This is important to assess if the positive effects that anti-malaria programs had on farmer productivity resulted in higher availability of natural resources per capita. This econometric strategy mimics the one I have used in Section 4, except that instead of considering farmer agricultural productivity, the dependent variable now consists in farm output per capita. This specification takes into account the increase that the eradication of malaria had on population in the US (Acemoglu et al., 2020; Gooch, 2017; Acemoglu and Johnson, 2007). The treatment is once again defined as the interaction between the exogenous county variation in weather conditions more or less suitable for the transmission of malaria disease and a post eradication indicator variable.

5.1 Econometric Specification

The regression presented below estimates the impact of the two indices measuring the prevalence of malaria (i.e. MSI and MEI) on the total county agricultural output per capita before and after the eradication of malaria in the US.

$$y_{i,s,t} = \alpha \ Malaria_i \cdot I_t^{Post} + \sum_{t=1870}^{2000} \Omega \mathbf{X'}_{i,s} I_t + \sum_{t=1870}^{2000} \delta_t I_t + \sum_p \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_q \omega_q I_t^q + \epsilon_{i,s,t} \tag{4}$$

This time the dependent variable $y_{i,s,t}$ is the natural log of the total farm output per capita as defined in section 3, in year t, county i and state s.

Similarly to equation 1 the variable $Malaria_i$ represent the historical incidence of malaria in county i before its eradication. $\mathbf{X}_{i,s}$ represents vectors of time invariant county-specific agricultural controls included in the regression. I include the same controls as in equation 1, i.e. county population density, the ratio of white population to county population, the ratio of farmland to total available county area, the proportion of farmers reporting the use of fertilizers, the proportion of farmland that reported have received drainage, the proportion of farmland with improved land and county average agriculture suitability. State fixed effects are represented as follows: $\sum_q \omega_q$. Where q indicates the set of US states. This time, the coefficient of interest α indicates if counties which had greater historical incidence of malaria in early 1900s, and whose agricultural productivity increased after the eradication of the vector borne disease, also had greater agricultural output per capita. With the inclusion of state fixed effects the coefficient of interest α is identified from within-state variation only, that is the empirical strategy compares counties within the same state.

5.2 Results

Results of equation 4 are reported in Table A12 and Table A13. They show that while counties with historical more suitable malarious weather conditions increased farmer agricultural productivity after the onset of malaria eradication campaigns they did not experience a statistically significant increment on the amount of agricultural output per capita. This last result emphasize the two-sided impact that the eradication of malaria had on the socio-economic development of the US namely. On one hand, the availability of an effective cure against malaria resulted in greater farmer agricultural productivity as a result of improved health conditions. On the other had, the eradication of malaria increased the population of more endemic counties, thus resulting in a non significant increase in the amount of agricultural output per capita. This is in line with what shown in Acemoglu and Johnson (2007) and Acemoglu et al. (2020) which focused on the effects pf medical innovations on population and economic development.

6 Concluding Remarks

This study has investigated the effects of medical innovations which brought to the successful eradication of malaria on farmer agricultural productivity growth in the US. The paper provides novel evidence that the eradication of malaria, occurred in early 1900s, resulted in an increase of farmer productivity in US counties with climatic conditions which made the transmission of vector-borne diseases more suitable. On the other hand, this study shows that the associated increase in population compensated the increase in agricultural output. As a result, the amount of agricultural output per capita did not not increase in more endemic counties compared to less endemic ones after the eradication of malaria.

I contribute to the literature investigating the effects of medical innovations on socio-economic development by showing that the eradication of a climate-related disease, such as malaria, besides increasing population, can also generate an increase in agricultural productivity. I then provide evidence that the increase in farmer agricultural productivity and farm output was caused by improved health conditions rather than a greater availability of arable land.

Second, economies largely dependent on agriculture and highly malarious are mostly located in Sub-Saharan Africa, Latin America and South-East Asia. The results shown in this paper could be useful to predict how these countries are economically affected by malaria and, more importantly, how their economies will respond to policies aimed at eradicating the vector-borne disease.

Overall, this paper provides novel insights at understanding the economics effects of medical innovation, motivating the example of the eradication of malaria as a disease which notably affects farmers living in rural areas resulting in positive agricultural productivity growth.

References

Acemoglu, D., Fergusson, L., and Johnson, S. (2020). Population and conflict. *The Review of Economic Studies*, 87(4):1565–1604.

Acemoglu, D. and Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of political Economy*, 115(6):925–985.

Alsan, M. (2015). The effect of the tsetse fly on african development. *American Economic Review*, 105(1):382–410.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Ashraf, Q. H., Lester, A., and Weil, D. N. (2008). When does improving health raise gdp? *NBER macroeconomics annual*, 23(1):157–204.

Bleakley, H. (2003). Disease and development: Evidence from the american south. *Journal of the European Economic Association*, 1(2-3):376–386.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. *The quarterly journal of economics*, 122(1):73–117.

Bleakley, H. (2009). Economic effects of childhood exposure to tropical disease. *American Economic Review*, 99(2):218–23.

Bleakley, H. (2010a). Health, human capital, and development. *Annu. Rev. Econ.*, 2(1):283–310.

Bleakley, H. (2010b). Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, 2(2):1–45.

Bleakley, H., Costa, D., and Lleras-Muney, A. (2014). Health, education, and income in the united states, 1820–2000. In *Human Capital in History: The American Record*, pages 121–159. University of Chicago Press.

Bleakley, H. and Hong, S. C. (2017). Adapting to the weather: lessons from us history. *The journal of economic history*, 77(3):756–795.

Brown, G. (1986). Control and eradication of malaria: Past, present and future.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., Stenlund, H., Martens, P., and Lloyd, S. J. (2014). Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, 111(9):3286–3291.

Cutler, D., Fung, W., Kremer, M., Singhal, M., and Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in india. *American Economic Journal: Applied Economics*, 2(2):72–94.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton University Press.

Derryberry, O. and Gartrell, F. (1952). Trends in malaria control program of the tennessee valley authority1. *The American journal of tropical medicine and hygiene*, 1(3):500–507.

Easterly, W. and Levine, R. (2016). The european origins of economic development. *Journal of Economic Growth*, 21(3):225–257.

EPA (1975). Ddt: A review of scientific and economic aspects of the decision to ban its use as a pesticide.

for Disease Control Prevention, U. C. (2010). The history of malaria, an ancient disease.

Gallup, J. L. and Sachs, J. D. (2001). The economic burden of malaria. *The American journal of tropical medicine and hygiene*, 64(1_suppl):85–96.

Giuliano, P. and Nunn, N. (2013). The transmission of democracy: From the village to the nation-state. *The American Economic Review*, 103(3):86.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5):165–70.

Gooch, E. (2017). The impact of reduced incidence of malaria and other mosquito-borne diseases on global population. *Journal of Development Economics*, 124:214–228.

Haines, M. R. et al. (2005). *Historical, demographic, economic, and social data: The United States, 1790-2002*. Inter-university Consortium for Political and Social Research.

Hansen, C. W. and Lønstrup, L. (2015). The rise in life expectancy and economic growth in the 20th century. *The Economic Journal*, 125(584):838–852.

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., and Snow, R. W. (2004). The global distribution and population at risk of malaria: past, present, and future. *The Lancet infectious diseases*, 4(6):327–336.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2017). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Humphreys, M. (2001). *Malaria: poverty, race, and public health in the United States*. JHU Press.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486–498.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E. (2017). New anthropogenic land use estimates for the holocene: Hyde 3.2. *Earth System Science Data*, 9(2):927–953.

Lysenko, A. and Semashko, I. (1968). Geography of malaria. a medico-geographic profile of an ancient disease. *Itogi Nauki: Medicinskaja Geografija*, pages 25–146.

Majori, G. (2012). Short history of malaria and its eradication in italy with short notes on the fight against the infection in the mediterranean basin. *Mediterranean journal of hematology and infectious diseases*, 4(1).

Medlock, J. M. and Leach, S. A. (2015). Effect of climate change on vector-borne disease risk in the uk. *The Lancet Infectious Diseases*, 15(6):721–730.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Newberry. Atlas of historical county borders. `https://publications.newberry.org/ahcbp/downloads/united_states.html`. Accessed: 2018-09-30.

Nunn, N. and Qian, N. (2011). The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics*, 126(2):593–650.

Percoco, M. (2013). The fight against disease: Malaria and economic development in italian regions. *Economic Geography*, 89(2):105–125.

Ramankutty, N., Foley, J. A., Norman, J., and McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography*, 11(5):377–392.

Robert, V., Macintyre, K., Keating, J., Trape, J.-F., Duchemin, J.-B., Warren, M., and Beier, J. C. (2003). Malaria transmission in urban sub-saharan africa. *The American journal of tropical medicine and hygiene*, 68(2):169–176.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek, M. (2015). Integrated public use microdata series: Version 6.0 [dataset]. minneapolis: University of minnesota.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680.

Saul, A. (2002). Malaria: Poverty, race, and public health in the united states.

Snowden, F. (2008). *The conquest of malaria: Italy, 1900-1962*. Yale University Press.

Tatem, A. J., Guerra, C. A., Kabaria, C. W., Noor, A. M., and Hay, S. I. (2008). Human population, urban settlement patterns and their impact on plasmodium falciparum malaria endemicity. *Malaria Journal*, 7(1):218.

Williams, R. C. (1952). The united states public health service, 1798–1950. *AJN The American Journal of Nursing*, 52(10):1279.

7 Figures and Tables

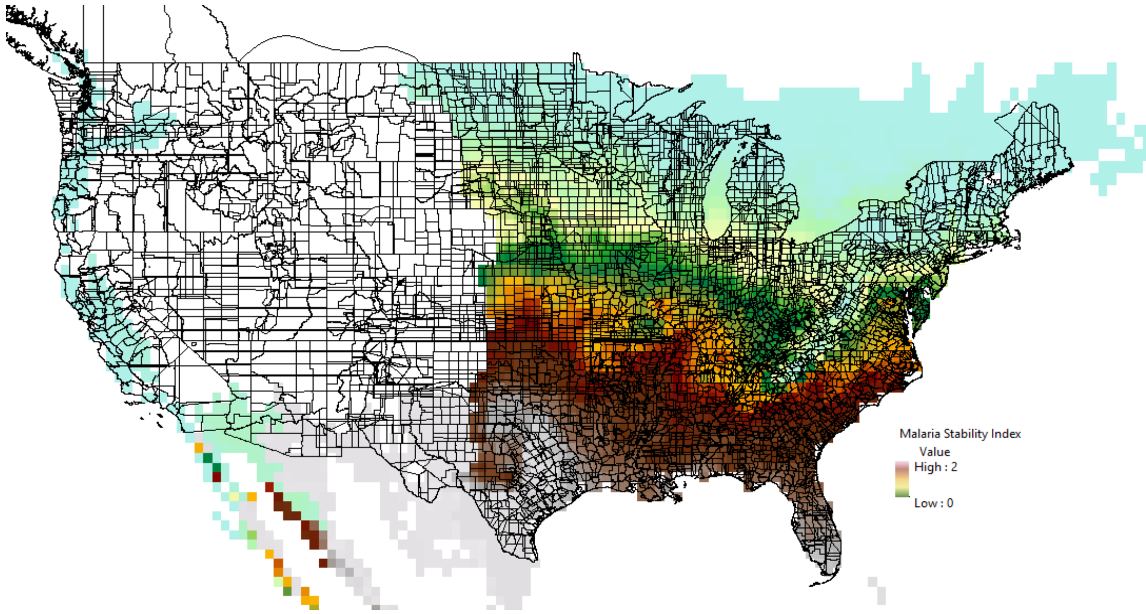


Figure 1: *The MSI captures the potential stability of malaria transmission based on regionally dominant vector mosquitoes, temperature and precipitation data set (Kiszewski et al., 2004). This is intersected with US historical county borders.*

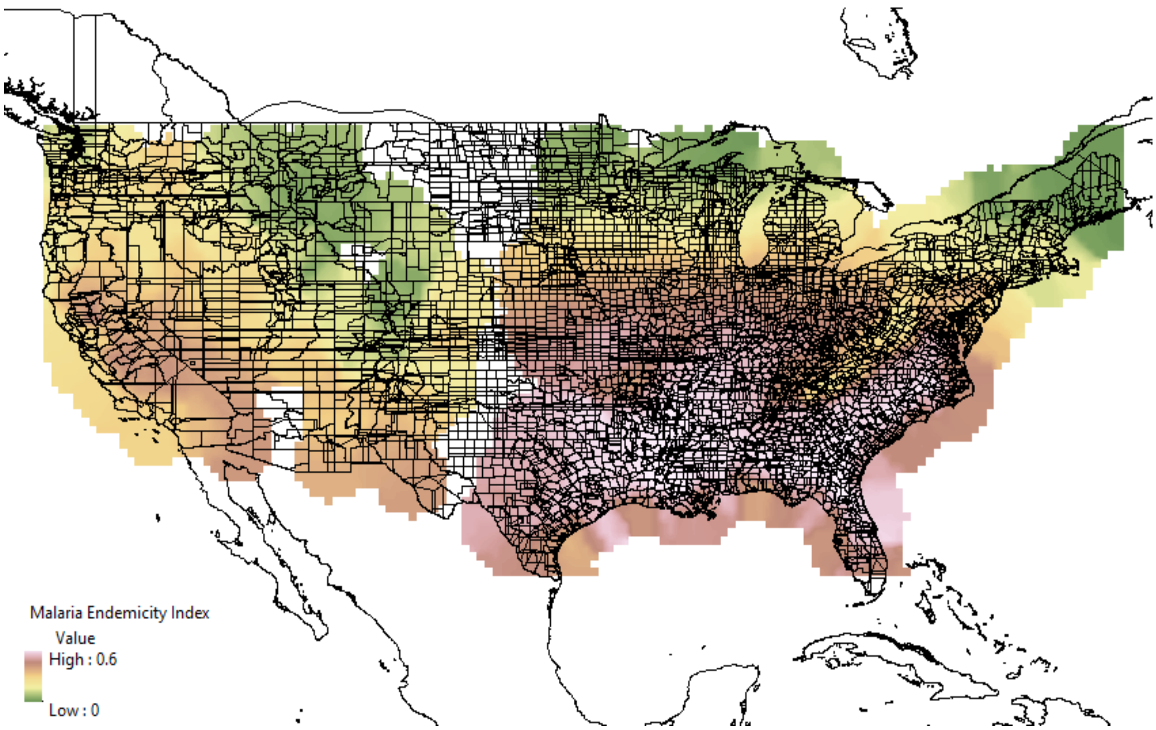


Figure 2: *The MEI captures the distribution of Malaria in the USA in 1900. Areas of high and low risk are merged to establish all-cause malaria transmission limits (Hay et al., 2004). This is intersected with historical US county borders.*

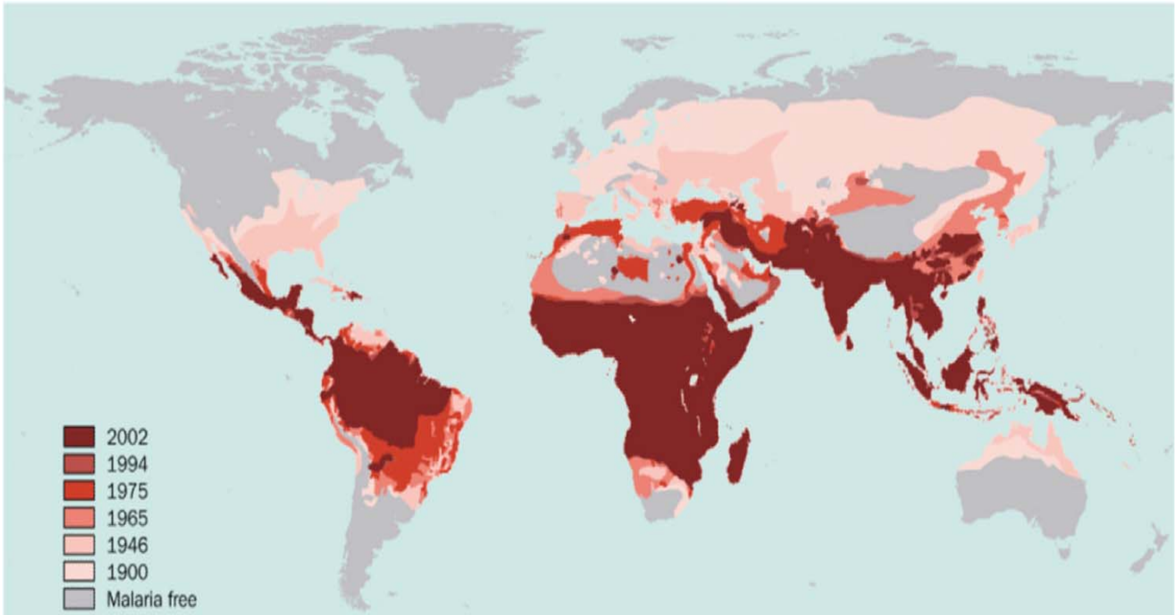


Figure 3: *Changes in the global distribution of malaria since 1900. Areas of high and low risk are merged to establish all-cause malaria transmission limits (Hay et al., 2004).*

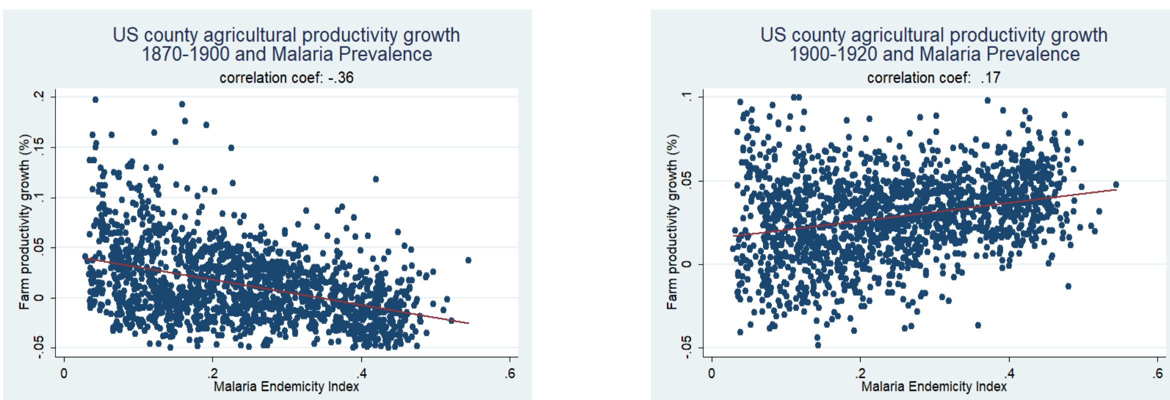


Figure 4: *Correlation between average annual agricultural productivity growth rates per US county and malaria prevalence in the US. Figure on the left shows agricultural productivity growth rates from 1870 to 1900, while figure on the right shows agricultural productivity growth rates from 1900 to 1920. This Figure shows that counties with higher prevalence of malaria had lower agricultural productivity per farmer compared to less endemic counties before the onset of the eradication program. The correlation reversed after the eradication campaigns started in the US. Source: Author's calculation using the geographical information system software ArcGis based on two historical records: Historical, Demographic and Social Data. The United States, 1790-2002" (Haines et al., 2005) and the IPUMS dataset (Ruggles et al., 2015).*

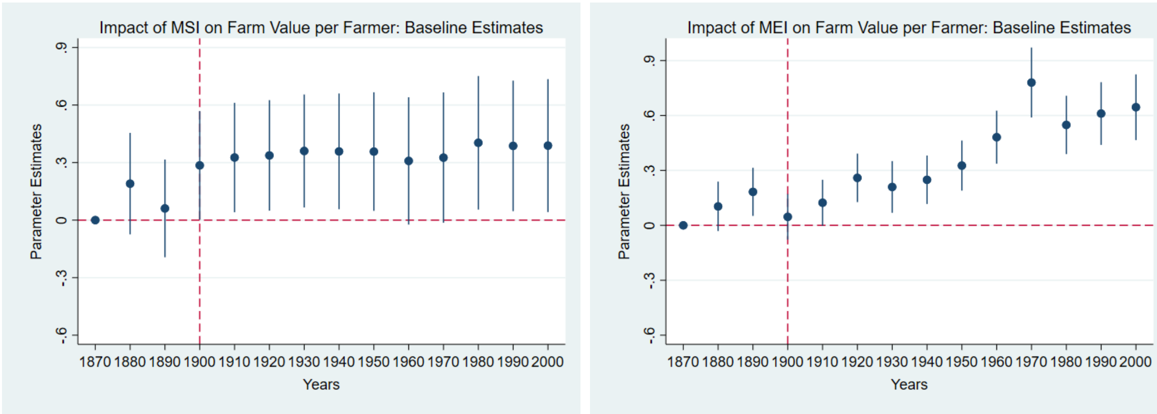


Figure 5: *Flexible coefficients of the relationship between farm value per farmer and MSI (left) and MEI (right): All counties.*

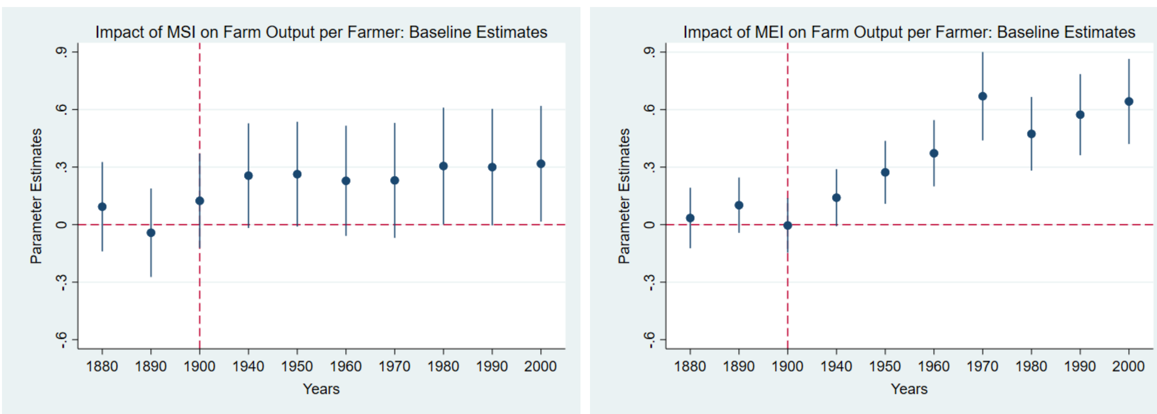


Figure 6: *Flexible coefficients of the relationship between farm output per farmer and MSI (left) and MEI (right): All counties.*

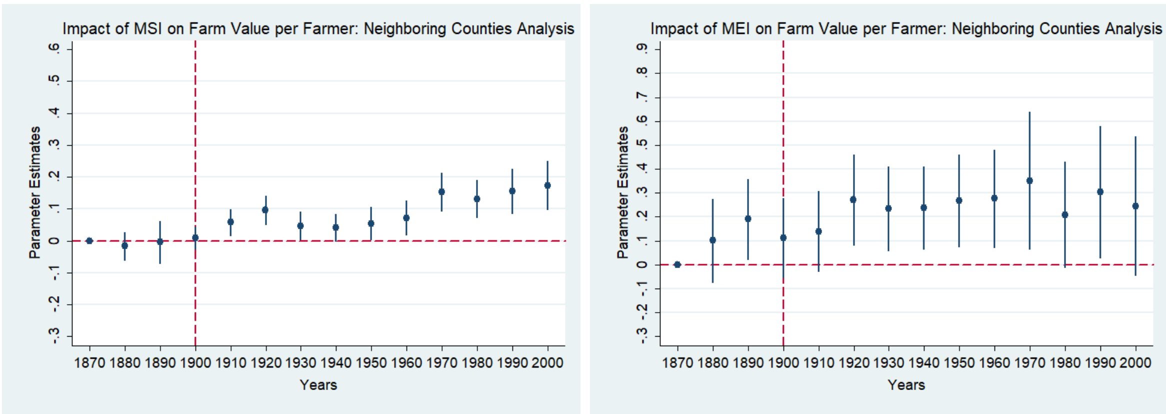


Figure 7: *Flexible coefficients of the relationship between farm value per farmer and MSI (left) and MEI (right): Neighboring Counties.*

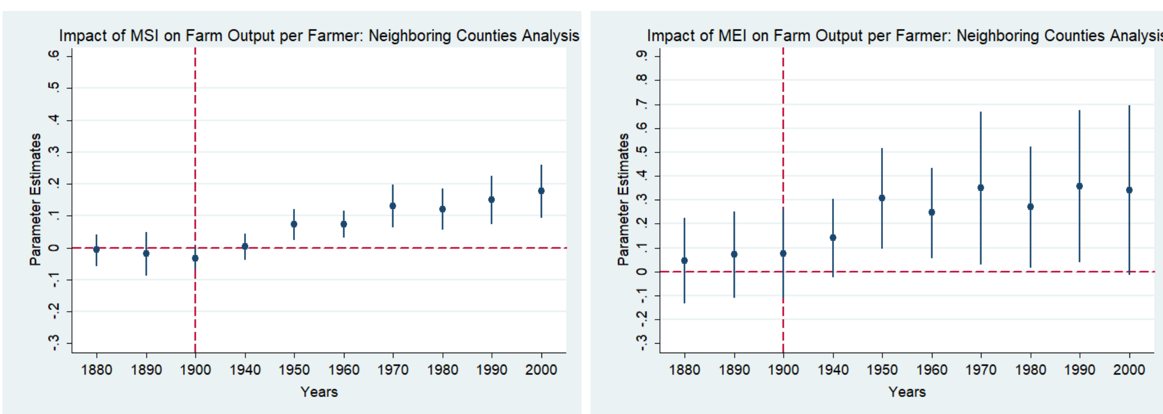


Figure 8: *Flexible coefficients of the relationship between farm output per farmer and MSI (left) and MEI (right): Neighboring Counties.*

Table 1: The impact of MSI on county agricultural productivity: Farm Value per Farmer

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MSI x Post	0.209***	0.196***	0.182***	0.181***
	(0.048)	(0.045)	(0.046)	(0.046)
Pop density		-8.188***	-5.614***	-5.603***
		(1.530)	(1.083)	(1.082)
White people		-0.833***	-1.096***	-1.095***
		(0.165)	(0.164)	(0.163)
Farmland			-1.288***	-1.286***
			(0.134)	(0.133)
Fertilizer			-0.999	-1.018
			(1.217)	(1.216)
Land improved			0.229	0.246
			(0.220)	(0.237)
Drainage			-0.000	0.010
			(0.091)	(0.094)
Agriculture suitability				-0.036
				(0.130)
State X Year	Yes	Yes	Yes	Yes
Observations	23783	23783	23769	23769
Adjusted R^2	0.778	0.783	0.788	0.788

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per farmer. The variable of interest is the average county Malaria Suitability Index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at county level, in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 2: The impact of MEI on county agricultural productivity: Farm value per farmer

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	0.142***	0.246***	0.265***	0.264***
	(0.040)	(0.044)	(0.047)	(0.047)
White people x MSI		-5.850***	-5.420***	-5.426***
		(1.536)	(1.561)	(1.569)
Pop density x MSI		-54.299***	-56.235**	-56.206**
		(20.654)	(21.945)	(21.961)
Farmland		-1.432***	-1.570***	-1.569***
		(0.110)	(0.134)	(0.132)
Fertilizer			-1.867	-1.871
			(1.198)	(1.197)
Land improved			0.362	0.366
			(0.221)	(0.237)
Drainage			-0.034	-0.032
			(0.091)	(0.093)
Agriculture suitability				-0.009
				(0.130)
State X Year	Yes	Yes	Yes	Yes
Observations	23783	23783	23769	23769
Adjusted R^2	0.778	0.789	0.789	0.789

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per farmer. The variable of interest is the average county Malaria Endemicity Index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at county level, in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 3: The impact of MSI on county agricultural productivity: Farm output per farmer

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MSI x Post	0.237***	0.225***	0.203***	0.211***
	(0.052)	(0.051)	(0.052)	(0.053)
Pop density		-8.825***	-7.156***	-7.228***
		(1.526)	(1.264)	(1.294)
White people		-0.845***	-1.013***	-1.017***
		(0.210)	(0.208)	(0.207)
Farmland			-1.311***	-1.327***
			(0.185)	(0.183)
Fertilizer			0.743	0.874
			(1.802)	(1.793)
Land improved			0.732**	0.621**
			(0.287)	(0.309)
Drainage			0.237**	0.169
			(0.117)	(0.120)
Agriculture suitability				0.243
				(0.160)
State X Year	Yes	Yes	Yes	Yes
Observations	18674	18674	18663	18663
Adjusted R^2	0.712	0.720	0.726	0.726

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm output per farmer. County total farm output data are missing for the period 1910-1940. The variable of interest is the average county Malaria Suitability Index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at county level, in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 4: The impact of MEI on county agricultural productivity: Farm output per farmer

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	0.160***	0.276***	0.300***	0.307***
	(0.053)	(0.059)	(0.062)	(0.062)
White people x MSI		-4.802**	-3.900*	-3.695*
		(2.017)	(2.034)	(2.037)
Pop density x MSI		-64.682**	-70.840**	-71.728**
		(25.294)	(28.763)	(29.042)
Farmland		-1.318***	-1.643***	-1.666***
		(0.142)	(0.185)	(0.183)
Fertilizer			-0.293	-0.173
			(1.747)	(1.739)
Land improved			0.865***	0.739**
			(0.291)	(0.311)
Drainage			0.205*	0.125
			(0.118)	(0.120)
Agriculture suitability				0.285*
				(0.159)
State X Year	Yes	Yes	Yes	Yes
Observations	18674	18674	18663	18663
Adjusted R^2	0.713	0.726	0.728	0.728

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm output per farmer. County total farm output data are missing for the period 1910-1940. The variable of interest is the average county Malaria Endemicity Index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at county level, in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 5: The impact of MSI on county agricultural productivity: Farm value per farmer. Neighboring Counties Analysis

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI x Post	0.103***	0.108***	0.099***	0.101***
	(0.027)	(0.026)	(0.028)	(0.026)
Farmland	-1.225***	-1.084***	-1.147***	-1.120***
	(0.118)	(0.122)	(0.131)	(0.131)
Pop density		-3.148***	-3.591***	-3.614***
		(0.672)	(0.849)	(0.775)
White people		0.085	0.053	0.147
		(0.110)	(0.143)	(0.155)
Fertilizer			1.639**	0.639
			(0.635)	(0.706)
Land improved			0.434*	0.722**
			(0.244)	(0.325)
Drainage			0.407***	0.506***
			(0.070)	(0.073)
Agriculture suitability				-0.413*
				(0.214)
State X Year	Yes	Yes	Yes	Yes
Observations	7375	7375	7361	7361
Adjusted R^2	0.868	0.870	0.872	0.873

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per farmer. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at the neighborhood level (i.e. 1°latitude x 1°longitude), in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 6: The impact of MEI on county agricultural productivity: Farm value per farmer. Neighboring Counties Analysis

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MEI x Post	0.164***	0.160***	0.090	0.100
	(0.061)	(0.060)	(0.060)	(0.066)
Farmland	-1.300***	-1.244***	-1.274***	-1.253***
	(0.121)	(0.119)	(0.134)	(0.130)
White people x MSI		1.358	0.537	1.944
		(1.551)	(2.118)	(2.146)
Pop density x MSI		-22.211***	-24.518***	-23.854***
		(5.610)	(7.018)	(6.943)
Fertilizer			1.081*	0.188
			(0.565)	(0.774)
Land improved			0.372	0.618*
			(0.246)	(0.321)
Drainage			0.382***	0.472***
			(0.075)	(0.073)
Agriculture suitability				-0.366*
				(0.213)
State X Year	Yes	Yes	Yes	Yes
Observations	7375	7375	7361	7361
Adjusted R^2	0.868	0.870	0.872	0.873

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm output per farmer. The variable of interest is the average county MEI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Coefficients are reported with standard errors, clustered at the neighborhood level (i.e. 1°latitude x 1°longitude), in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 7: The impact of MSI on county farm value per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900	1870-1900 Post= 1890, 1900	1890-1920 Post=1900, 1910, 1920	1890-1920 Post=1910, 1920	1940-1970 Post= 1950, 1960 1970	1940-1970 Post= 1960, 1970
	b/se	b/se	b/se	b/se	b/se	b/se
MSI x Post	0.174	0.039	0.258***	0.166***	0.004	-0.013
	(0.111)	(0.028)	(0.085)	(0.028)	(0.022)	(0.028)
State X Year	Yes	Yes	Yes	Yes	Yes	Yes
Farmland	Yes	Yes	Yes	Yes	Yes	Yes
Pop density	Yes	Yes	Yes	Yes	Yes	Yes
White people	Yes	Yes	Yes	Yes	Yes	Yes
Fertilizer	Yes	Yes	Yes	Yes	Yes	Yes
Drainage	Yes	Yes	Yes	Yes	Yes	Yes
Land improved	Yes	Yes	Yes	Yes	Yes	Yes
Agriculture suitability	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808	6808	6808
Adjusted R^2	0.863	0.856	0.880	0.877	0.817	0.821

Notes: Observations are at the county-decade level. All regressions use a baseline sample of 1700 Old World countries. The periods vary by specification and are reported in the column headings. The dependent variable is the natural log of farm value per farmer. MSI indicates the average county Malaria Suitability Index. The Post indicator variable varies by specification. Its definition is reported in the column headings. All regressions include year fixed effects, county fixed effects, and each of the following baseline county specific controls interacted with the full set of time-period fixed effects: Proportion of farmland; Population density; proportion of white people; Use of fertilizers; Proportion of land using methods of drainage; Agricultural Suitability Index. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at county level, in parentheses.

8 Appendix

Table 8: Summary Statistics: Farm Productivity Measures and Population, Weather and Agricultural controls

	mean	sd	min	max
Ln fv/farmer	10.013	1.679	-5.842	23.983
Ln fout/farmer	8.502	1.425	-6.164	22.211
Ln fv/acre	5.609	1.532	-4.824	9.169
Ln fout/acre	4.092	1.354	-4.523	7.418
MSI	0.055	0.065	0.000	1.930
MEI	0.244	0.123	0.029	0.545
Farmland	0.489	0.301	0.000	1.000
Pop density	0.027	0.035	0.000	0.567
White people	0.839	0.217	0.108	1.000
Fertilizer	0.012	0.024	0.000	0.187
Drainage	0.149	0.274	0.000	1.000
Land improved	0.498	0.225	0.004	0.971
Agr. Suitability	0.525	0.335	0.007	0.998

Source: Author's calculations based on two historical records: Historical, Demographic and Social Data:" "The United States, 1790-2002 (Haines et al., 2005) and the IPUMS dataset (Ruggles et al., 2015)

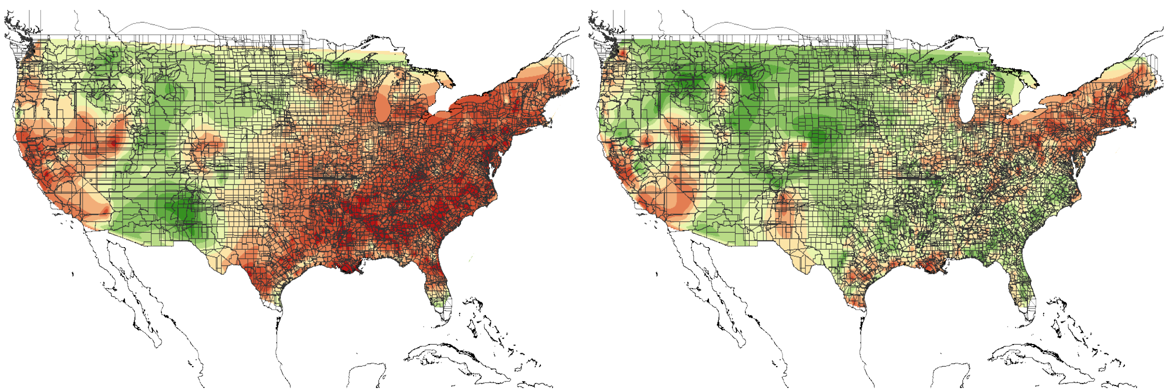


Figure 9: *Average annual agricultural productivity growth rates per US county from 1870 to 1900 (left) and from 1900 to 1920 (right). Source: Author's calculation using the geographical information system software ArcGis based on two historical records: Historical, Demographic and Social Data. The United States, 1790-2002" (Haines et al., 2005) and the IPUMS dataset (Ruggles et al., 2015).*

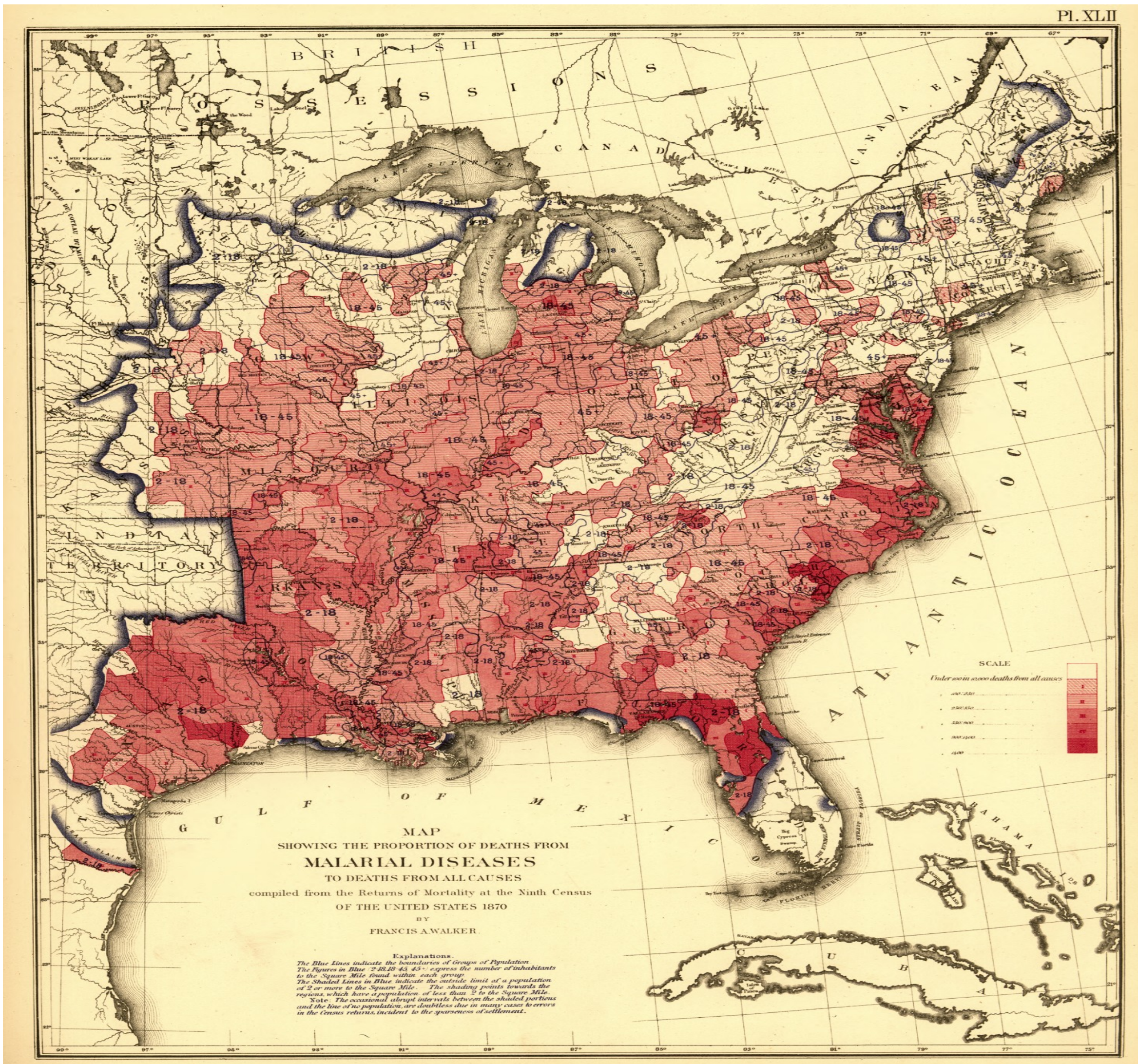


Figure 10: *Proportion of deaths caused by malaria to deaths from all causes. Map retrieved from the Statistical Atlas from the* 9^{th} *Census of the United States 1870.*

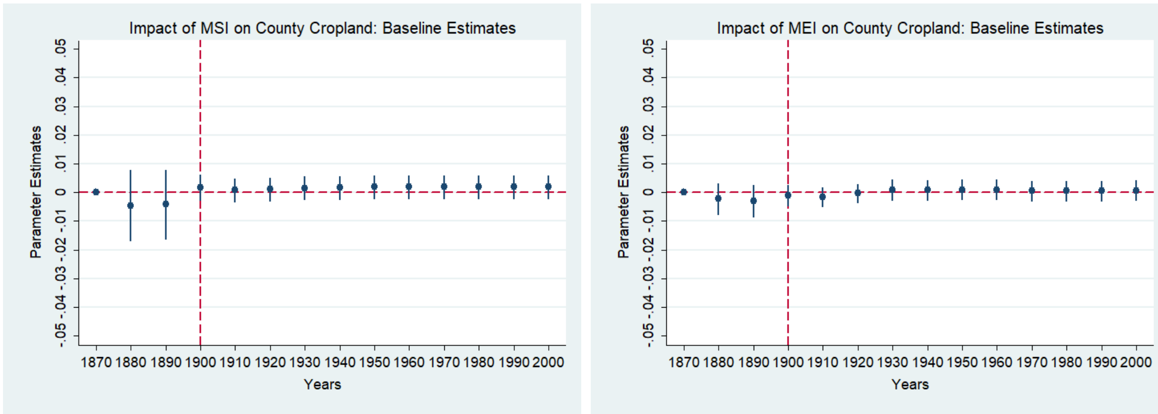


Figure 11: *Flexible coefficients of the relationship between proportion of cropland out of total county area and MSI (left) and MEI (right): All counties.*

8.1 Correlation between malaria endemicity/stability measures and control variables

Table 9: Correlation Table for each measure of malaria

	MEI	MSI
	b	b
Ln fout/farmer	0.005	0.003
MSI	0.629	
Pop density	-0.157	-0.043
Farmland	0.082	-0.013
White people	-0.249	-0.102
Fertilizer	0.068	0.048
Drainage	0.015	-0.014
Land improved	-0.160	-0.069
Agr. Suitability	0.040	0.029
Constant	0.399	0.145
Observations	18663	18663

Table 9 shows correlation between control variables and dependent with the two measures of malaria to check for the presence of multicollinearity. The primary concern is that as the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get wildly inflated.

To check for multicollinearity I compute the variance inflation factor (VIF). As a rule of thumb, a variable whose VIF value is greater than 10 may merit further investigation. It means that the variable could be considered as a linear combination of other independent variables. The regression model predicting the natural logarithm of farm value and farm

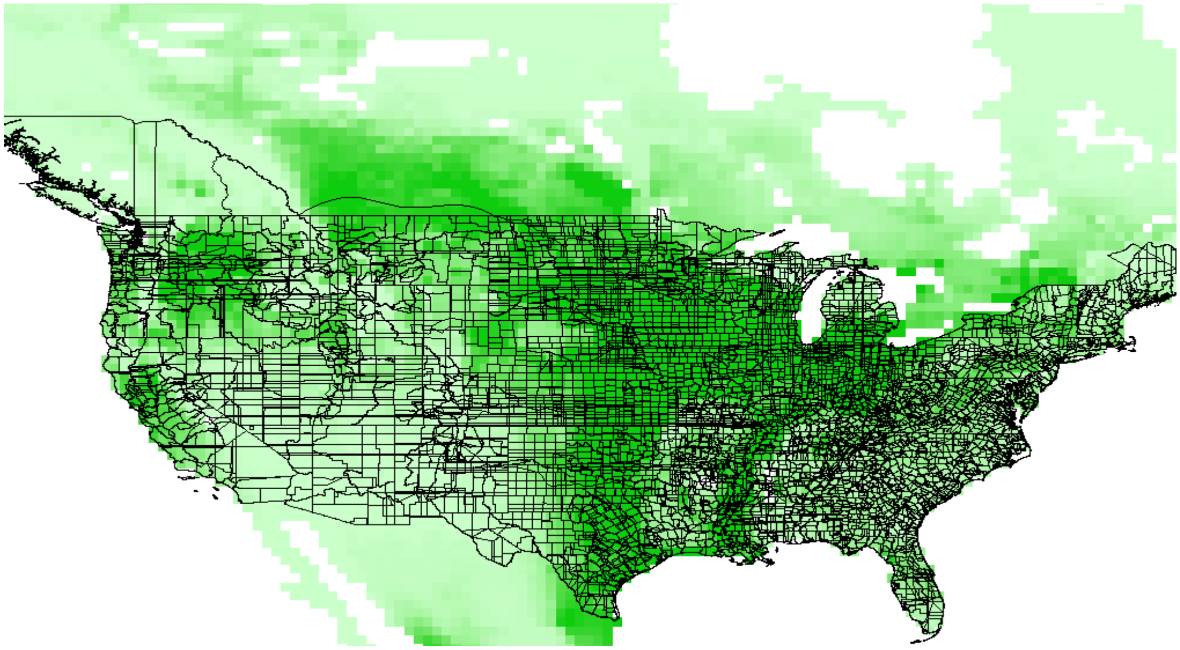


Figure 12: *Distribution of the total crop suitability per county (Ramankutty et al., 2002).*

output per farmer from the MSI population density in 1870, proportion of farmland out of total county acre in 1870, the proportion of white people out of total number of farmers in 1870, the proportion of farmers reporting the use of fertilizers, have received drainage and have improved land in 1870, and agriculture suitability index is shown with the associated VIFs in Table 15. VIFs exclude any problems of multicollinearity in the model, thus the empirical model presented in equation 1 gives valid results about the predictor and the control variables.

Table 10: Check for multicollinearity of predictors

	Ln fv/farmer VIF
MSI	1.25
Pop density	1.45
Farmland	1.54
White people	1.48
Fertilizer	1.31
Drainage	1.26
Land improved	1.97
Agr. Suitability	1.82
Constant	
Mean VIF	1.53

Table 11: Check for multicollinearity of predictors

	(1) Ln farmers vif
MEI	2.571945
MSI	1.54874
Pop density	1.457751
Farmland	1.658303
White people	1.989921
Fertilizer	1.388199
Drainage	1.272055
Land improved	2.176409
Agr. Suitability	1.868958
Constant	
Observations	23769

Table 12: The impact of MSI on county agricultural output per capita

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MSI x Post	-0.023	-0.017	-0.005	-0.005
	(0.033)	(0.036)	(0.031)	(0.032)
White people		0.275*	0.204	0.204
		(0.141)	(0.141)	(0.141)
Fertilizer			-1.600	-1.605
			(1.313)	(1.310)
Land improved			-0.679***	-0.675***
			(0.167)	(0.177)
Drainage			-0.102	-0.100
			(0.075)	(0.080)
Agriculture suitability				-0.008
				(0.105)
State X Year	Yes	Yes	Yes	Yes
Observations	18733	18733	18722	18722
Adjusted R^2	0.737	0.737	0.739	0.739

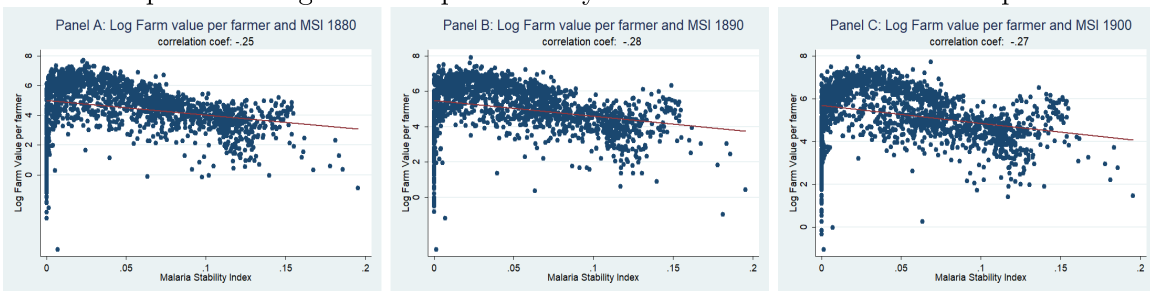
Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm output per capita. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Coefficients are reported with standard errors, clustered at the neighborhood level (i.e. 1°latitude x 1°longitude), in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 13: The impact of MEI on county agricultural output per capita

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	0.052	0.083**	0.075*	0.075*
	(0.036)	(0.039)	(0.040)	(0.040)
White people		0.394***	0.311**	0.311**
		(0.148)	(0.149)	(0.149)
Fertilizer			-1.915	-1.912
			(1.319)	(1.316)
Land improved			-0.645***	-0.648***
			(0.168)	(0.178)
Drainage			-0.115	-0.116
			(0.076)	(0.080)
Agriculture suitability				0.005
				(0.104)
State X Year	Yes	Yes	Yes	Yes
Observations	18733	18733	18722	18722
Adjusted R^2	0.737	0.737	0.740	0.740

Notes: The periods range from 1870 to 2000, and are observed every other decade.The dependent variable is the natural log of the county average farm output per capita. The variable of interest is the average county MEI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Coefficients are reported with standard errors, clustered at the neighborhood level (i.e. 1°latitude x 1°longitude), in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Relationship between agricultural productivity and MSI *before* vector control policies



Relationship between agricultural productivity and MSI *after* vector control policies

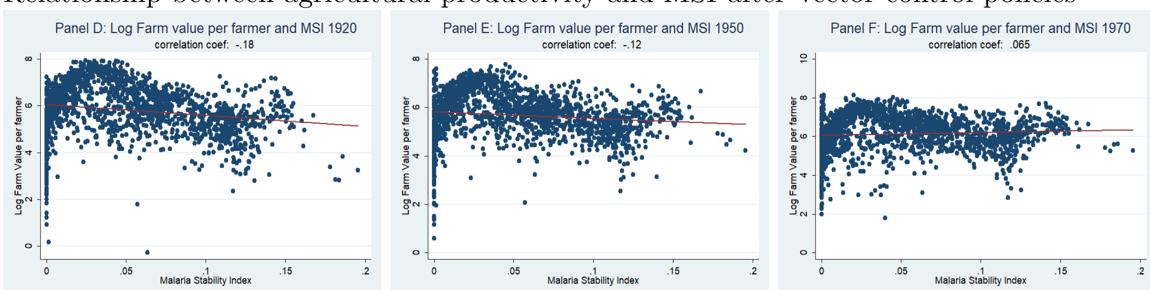
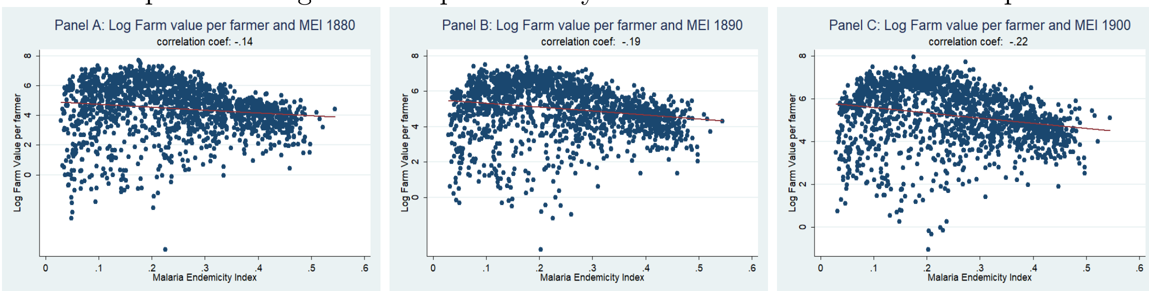


Figure 13: *Correlation of malaria stability levels and log county farm value per farmer in 1880, 1890, 1900, 1920, 1950 and 1970.*

Relationship between agricultural productivity and MEI *before* vector control policies



Relationship between agricultural productivity and MEI *after* vector control policies

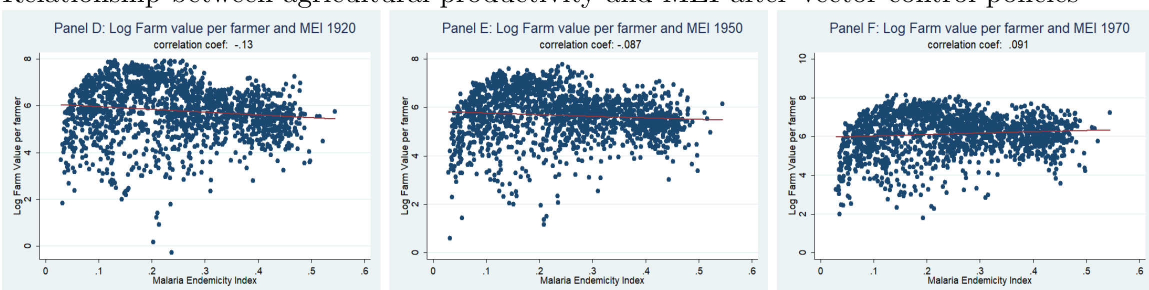


Figure 14: *Correlation of malaria stability levels and log county farm value per farmer in 1880, 1890, 1900, 1920, 1950 and 1970.*

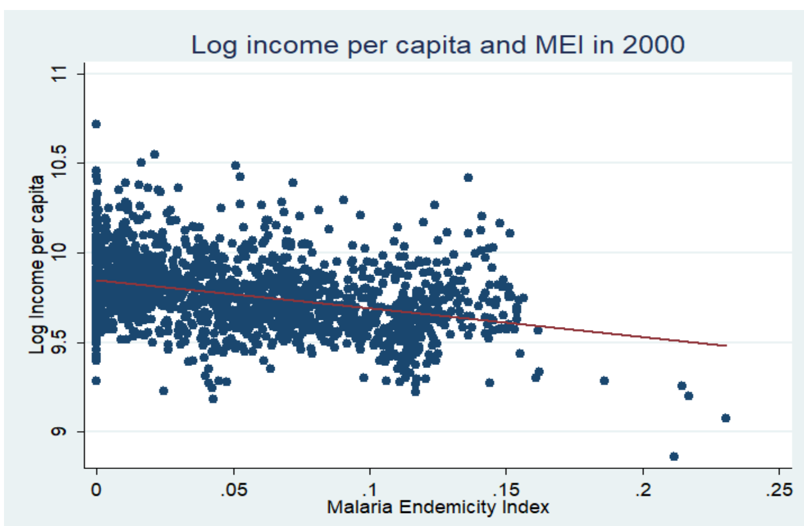


Figure 15: *Correlation between malaria stability index and log income per capita per county in 2000.*

8.2 Robustness

Table 14: The impact of MEI on county farm value per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900	1870-1900 Post= 1890, 1900	1890-1920 Post=1900, 1910, 1920	1890-1920 Post=1910, 1920	1940-1970 Post= 1950, 1960 1970	1940-1970 Post= 1960, 1970
	b/se	b/se	b/se	b/se	b/se	b/se
MEI x Post	-0.004	-0.044	-0.083**	0.059**	0.116***	0.163***
	(0.057)	(0.044)	(0.038)	(0.026)	(0.029)	(0.036)
State X Year	Yes	Yes	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
MSI	Yes	Yes	Yes	Yes	Yes	Yes
Ln(pop den) X MSI	Yes	Yes	Yes	Yes	Yes	Yes
Ln(p white) X MSI	Yes	Yes	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808	6808	6808
Adjusted R^2	0.864	0.857	0.881	0.877	0.818	0.822

*Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

Table 15: The impact of MSI on county amount of cropland

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MSI x Post	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Farmland	-0.004*	-0.004*	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Pop density		0.007	0.014	0.013
		(0.009)	(0.011)	(0.011)
White people		0.001	0.003	0.003
		(0.004)	(0.005)	(0.005)
Fertilizer			-0.003	-0.008
			(0.024)	(0.025)
Land improved			-0.005	-0.004
			(0.005)	(0.006)
Drainage			0.000	0.001
			(0.001)	(0.001)
Agriculture suitability				-0.001
				(0.003)
State X Year	Yes	Yes	Yes	Yes
Observations	23776	23776	23763	23763
Adjusted R^2	0.999	0.999	0.999	0.999

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county hectares of cropland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1940 to 2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Full details of each control variable are provided in the text and the Appendix.Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 16: The impact of MEI on county amount of cropland

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	-0.001	-0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Farmland	-0.003	-0.003	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
White people x MSI		-0.011	-0.002	-0.001
		(0.027)	(0.027)	(0.028)
Pop density x MSI		-0.152	-0.100	-0.096
		(0.110)	(0.095)	(0.088)
Fertilizer			0.000	-0.002
			(0.022)	(0.023)
Land improved			-0.006	-0.006
			(0.004)	(0.006)
Drainage			0.001	0.001
			(0.001)	(0.001)
Agriculture suitability				-0.001
				(0.003)
State X Year	Yes	Yes	Yes	Yes
Observations	23776	23776	23763	23763
Adjusted R^2	0.999	0.999	0.999	0.999

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county hectares of cropland. The variable of interest is the average county MEI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1940 to 2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Full details of each control variable are provided in the text and the Appendix.Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 17: The impact of MSI on county agricultural productivity: Farm output per farmer. Neighboring Counties Analysis

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MSI x Post	0.132***	0.137***	0.115***	0.119***
	(0.035)	(0.033)	(0.033)	(0.030)
Farmland	-1.242***	-1.109***	-1.337***	-1.293***
	(0.187)	(0.206)	(0.214)	(0.207)
Pop density		-2.913***	-4.134***	-4.175***
		(0.785)	(1.067)	(0.937)
White people		0.094	0.004	0.162
		(0.188)	(0.241)	(0.247)
Fertilizer			2.660**	0.983
			(1.260)	(1.281)
Land improved			1.336***	1.819***
			(0.349)	(0.427)
Drainage			0.473***	0.639***
			(0.085)	(0.097)
Agriculture suitability				-0.692**
				(0.269)
State X Year	Yes	Yes	Yes	Yes
Observations	5794	5794	5783	5783
Adjusted R^2	0.747	0.750	0.763	0.766

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm output per farmer. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1940 to 2000. Farm output per US county are missing from 1910 to 1940. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at the neighborhood level (i.e. 1°latitude x 1°longitude), in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 18: The impact of MEI on county agricultural productivity: Farm output per farmer. Neighboring Counties Analysis

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	0.164*	0.170**	0.092	0.111
	(0.086)	(0.085)	(0.083)	(0.089)
Farmland	-1.317***	-1.226***	-1.447***	-1.409***
	(0.196)	(0.205)	(0.221)	(0.211)
White people x MSI		3.916	2.362	4.906
		(2.874)	(3.979)	(3.993)
Pop density x MSI		-25.326***	-31.708***	-30.511***
		(7.303)	(10.753)	(10.573)
Fertilizer			2.490**	0.875
			(1.199)	(1.293)
Land improved			1.263***	1.709***
			(0.354)	(0.426)
Drainage			0.449***	0.610***
			(0.087)	(0.096)
Agriculture suitability				-0.661**
				(0.266)
State X Year	Yes	Yes	Yes	Yes
Observations	5794	5794	5783	5783
Adjusted R^2	0.748	0.752	0.764	0.766

Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm output per farmer. The variable of interest is the average county MEI index. The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1940 to 2000. Farm output per US county are missing from 1910 to 1940. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes. Full details of each control variable are provided in the text and the Appendix. Coefficients are reported with standard errors, clustered at the neighborhood level (i.e. 1°latitude x 1°longitude), in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 19: The impact of MSI on county farm output per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900	1870-1900 Post= 1890, 1900	1940-1970 Post= 1950, 1960 1970	1940-1970 Post= 1960, 1970
	b/se	b/se	b/se	b/se
MSI x Post	0.043	-0.011	0.018	-0.002
	(0.097)	(0.028)	(0.030)	(0.021)
State X Year	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes
Popden 1870 X Year	Yes	Yes	Yes	Yes
Pwhite 1870 X Year	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808
Adjusted R^2	0.808	0.803	0.804	0.806

*Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

Table 20: The impact of MEI on county farm output per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900	1870-1900 Post= 1890, 1900	1940-1970 Post= 1950, 1960 1970	1940-1970 Post= 1960, 1970
	b/se	b/se	b/se	b/se
MEI x Post	-0.030	-0.062	0.146***	0.147***
	(0.053)	(0.042)	(0.034)	(0.040)
State X Year	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes
MSI	Yes	Yes	Yes	Yes
Ln(pop den) X MSI	Yes	Yes	Yes	Yes
Ln(p white) X MSI	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808
Adjusted R^2	0.809	0.804	0.804	0.805

*Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

Robustness check 4: Discrete measure of the Malaria Endemicity Index. As in Kiszewski et al. (2004) the MSI is divided into 9 categories ranging from malaria free grids with a MSI between 0 and 0.05 to the highest persistence of malarious weather conditions with a MSI between 26 and 39[28]:

Table 21: The impact of malaria eradication on county agricultural productivity: Baseline estimates

	ln(county average farm value per acre)		
	b/se	b/se	b/se
MEI	0.168***	0.231***	0.200***
	(0.056)	(0.056)	(0.055)
State X Year	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes
Popden 1870 X Year	No	Yes	Yes
Pwhite 1870 X Year	No	Yes	Yes
Farmer 1870 X Year	No	No	Yes
Fertilizer 1870 X Year	No	No	Yes
Suitability X Year	No	No	Yes
Observations	23842	23842	23842
Adjusted R^2	0.939	0.941	0.943

Notes: The periods range from 1870 to 2000, and are observed every other decade.
The dependent variable is the natural log of the county average farm value per county acre of farmland.
The variable of interest is the average county MEI index.
The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000.
All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects.
The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes.
In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

[28]Since the MSI for the USA predominantly ranges from 0 to 1, the table shows the effect of eradicating malaria on the agricultural productivity of counties being in the second category compared to the counties belonging in the first category

GREEN
Centre for Geography, Resources, Environment, Energy and Networks
via Röntgen, 1
20136 Milano - Italia

www.green.unibocconi.eu