Malaria in the Americas:
A Retrospective Analysis of Childhood Exposure*

Hoyt Bleakley†

June 16, 2006

Abstract
This study considers the malaria-eradication campaigns in the United States (circa 1920), and in Brazil, Colombia and Mexico (circa 1955), with a specific goal of measuring how much childhood exposure to malaria depresses labor productivity. The eradication campaigns studied happened because of advances in medical and public-health knowledge, which mitigates concerns about reverse causality of the timing of eradication efforts. Data from regional malaria eradication programs are collected and collated with publicly available census data. Malarious areas saw large drops in their malaria incidence following the campaign. In both absolute terms and relative to non-malarious areas, cohorts born after eradication had higher income as adults than the preceding generation. Similar increases in literacy and the returns to schooling are observed. Results for years of schooling are mixed. An analysis at the year-of-birth level indicates that the observed changes coincide with childhood exposure to the campaigns rather than to a pre-existing trend.

Keywords: Malaria, returns to health.
JEL codes: I12, J24, 010, H43.

∗Preliminary and incomplete. This study subsumes two earlier papers “Malaria and Human Capital: Evidence from the American South” and “Malaria Eradication in Colombia and Mexico: A Long-Term Follow-up.” The author gratefully acknowledges the contribution of many in preparing this study. Funding support came from the University of California Pacific Rim Research Program, the UCSD Faculty Senate’s Committee on Research, the UCSD Hellman Faculty Fellowship, and the Graduate School of Business of the University of Chicago. This paper was partially prepared while the author was visiting the Center for US/Mexican Studies at UCSD and the Universidad de los Andes. Imelda Flores, Dr. Mauricio Vera, and Dr. Victor Orlano of Colombia’s Instituto Nacional de Salud provided useful guidance in interpreting the Colombian malaria data. Glenn Hyman of the Centro Internacional de Agricultura Tropical shared data on the Colombian municipio boundaries. Andrew Mellinger provided the malaria-ecology raster data. Librarians too numerous to mention from UC-San Diego, la Universidad Nacional de Colombia, la Fundación Santa Fe de Bogotá, and the University of Chicago have consistently gone the extra mile in aiding my search for data. I have also benefited from the comments of Jennifer Baca, Eli Berman, Mónica García, Jonathan Guryan, Gordon Hanson, Moramay López, Adrienne Lucas, Paola Mejía, Emilio Quevedo, Rocio Ribero, Fabio Sanchez, Duncan Thomas, Carol Vargas, David Weil, and seminar participants at UC-San Diego’s Center for US/Mexican Studies, Universidad de los Andes, Yale University, the University of Chicago, Columbia University, MIT, the NBER Cohort Studies group, and Princeton University. Jennifer Baca, Barbara Cunha, Rebeca Mohr, Michael Pisa, Tareq Rashidi, Lisandra Rickards, and Raghavan Selvara all provided able research assistance.

†Assistant Professor of Economics, Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL, 60637. Telephone: (773) 834-2192. Electronic mail: bleakley[at]chicagogsb[dot]edu
1 Introduction

The disease known as malaria, a “scourge of mankind” through history, persists in tropical countries up to the present day. These same tropical areas have, generally speaking, a much lower level of economic development than that enjoyed in the temperate climates. These facts lead us to a natural question: does malaria hold back economic progress?

The simple correlation between tropical disease and productivity cannot answer this question. Malaria might depress productivity, but the failure to eradicate malaria might equally well be a symptom of underdevelopment. Indeed, tropical countries also tend to have “debilitating” institutions, such as the poor protection of property rights and weak rule of law, the latter of which makes it difficult to marshal resources in support of public health. This important international question has an interesting parallel among regions within countries. For example, southern Mexico, the southern U.S., the tierra caliente of Colombia, and the north of Brazil have borne a disproportionate burden of malaria infection in those countries, but these regions were also disproportionately host to colonial, extractive institutions for several centuries. Both factors plausibly play a role in the failure to eradicate malaria.

How can we cut through this Gordian knot of circular causality? The standard econometric answer is to consider plausibly exogenous variation in malaria. A possible source of such variation comes from targeted interventions in public health.

The present study considers two major attempts to eradicate malaria in the Americas during the Twentieth century. The first episode analyzed took place in the Southern United States, largely in the decade of the 1920s. During this period, a wealth of new knowledge about the malaria transmission mechanism was applied to the malaria problem in the South. The second episode is the worldwide malaria eradication campaign, and in particular as it was implemented in Colombia, Brazil, and Mexico (principally in the 1950s). The efforts to eradicate malaria worldwide were spurred on by the discovery of DDT, a powerful pesticide. After the World War II, the World Health Organization helped many afflicted countries put together programs of spraying to combat malaria transmission. The campaigns in these regions partially interrupted the malaria transmission
cycle and brought about marked drops in infection in a relatively short period of time. (Further background on the disease and the eradication efforts is found in Section 2.) Additionally, sufficient time has passed that we can evaluate the long-term consequences of eradication.

The relatively rapid impact of the treatment campaigns combine with cross-area heterogeneity to form the research design of the present study. These four countries are geographically variegated, such that, within each country, some regions have climates that support malaria transmissions, while other regions do not. Areas with high malaria infection rates had more to gain from eradication, but the non-malarious areas serve as a comparison group, filtering out common trends in national policy, for example. Moreover, the reductions in disease burden occur in the space of a few years, and resulted from critical innovations to knowledge and spending, and these innovations came largely from outside the studied areas. This latter fact mitigates the usual concern about policy endogeneity.

A further goal of this study is identify the role that childhood exposure to malaria plays in subsequent labor productivity as an adult. While direct effects of malaria on adults can be partially measured with lost wages from work absences and mortality, little is known about effects that persist over the life cycle. Children are more susceptible to malaria than adults, most likely because prolonged exposure to the disease brings some degree of resistance. But while partial immunity is conferred by age, the damage from childhood exposure to malaria may be hard to undo: most of a person’s human-capital and physiological development happens in childhood. On the physiological side, a malaria-free childhood might mean that the individual is more robust as an adult, with concomitant increases in labor supply. On the human-capital side, more oxygen getting to the brain translates into more learning. This would be manifested in the data as greater literacy, higher adult earnings, and, for a fixed time in school, higher returns to schooling. This might also affect the schooling decision, but, because malaria also affects the childhood wage (the opportunity cost of schooling), this latter effect is ambiguously signed by the theory. Malaria’s possible effect contemporaneously on wages implies that an additional channel is via parental income.

I show in Section 5 that childhood exposure to malaria is indeed related to lower income as an adult. Using census microdata, I compare the socioeconomic outcomes of cohorts born well before
the campaigns to those born afterwards. In both absolute terms and relative to the comparison group of non-malarious areas, cohorts born after eradication had higher income and were more literate. Mixed results are found for years of schooling, consistent with the economic theory of schooling.

This result is not sensitive to accounting for a variety of alternative hypotheses. I obtain essentially similar estimates of malaria coefficients even when controlling for several indicators of health and economic development. Moreover, I show in Section 4 that the shift in the malaria-income relationship coincides with childhood exposure to the eradication efforts, and not with a trend or autoregressive process. I also find a relative increase in the returns to schooling associated with malaria eradication.

2 Malaria and the Eradication Campaigns

2.1 Malaria: Symptoms and History of the Disease

Malaria is a parasitic disease that afflicts humans. The parasite is a protozoan of the genus *Plasmodium* and has a complicated life-cycle that is partly spent in a mosquito “vector” and partly in the human host’s bloodstream. The disease is transmitted when a mosquito takes a blood meal from an infected person and, some time later, bites another person. Acute symptoms of infection include fever and shivering. The main chronic symptom is anemia. Malaria results in death on occasion, but the strains prevalent in the Americas (vivax and malariae) have low case-fatality rates compared to the predominantly African variety (falciparum).

Malaria has been present throughout recorded history. As recently as a few centuries ago, it extended into areas such as Northern European, where in England it was known as the “ague.” By the turn of the 20th century, malaria had gradually receded to tropical and subtropical regions.

The turn of the 20th century saw considerable advances in the scientific understanding of the disease. Doctor Charles Louis Alphonse Laveran, of the French army, showed in the early 1890s through microscopic studies that malaria is caused by a protozoan. Dr. (later Sir) Ronald Ross, of the British Indian Medical Service, discovered in the late 1890s that malaria is transmitted via
mosquitoes. Both men later won the Nobel Prize for Medicine.

2.2 Efforts against Malaria in the Southern U.S., circa 1920

The US government’s interest in vector-borne diseases arose in the 20th century not because of a new-found interest in the Southern region, but because of the acquisition of Cuba and of the Panama Canal Zone. Early in the occupation of Cuba, the US Army dispatched a team of physicians, among them Dr. Walter Reed, to Havana to combat yellow fever and malaria. Armed with the new knowledge about these diseases, the Army was able to bring these diseases under control in that city. Another team of American physicians, this time led by Dr. William Gorgas, were able to bring these diseases under control in the Zone, which was a considerable challenge given that much of the area was a humid, tropical jungle.¹

The progress made by US Army doctors against malaria in Cuba and Panama inspired work back home in the South in the latter half of the 1910s. Several physicians in the United States Public Health Service (PHS) began collecting information on the distribution of malaria throughout the South and the prevalence of the various species of parasites and mosquitoes.² The PHS began actual treatment campaigns in a limited way, first by controlling malaria in a handful of mill villages (to which the Service had been invited by the mill owners). The Rockefeller Foundation, having mounted a successful campaign against hookworm in the early 1910s, also funded anti-malarial work through its International Health Board (IHB). These two groups sponsored demonstration projects in a number of small, rural towns across the South. They employed a variety of methods

¹It is doubtful that the construction of the Canal would have been economically feasible were it not for these sizeable innovations to knowledge. The following anecdote is illustrative of the primitive state of medical knowledge of malaria just a few years earlier:

And all the while, in the lovely gardens surrounding the hospital, thousands of ring-shaped pottery dishes filled with water to protect plants and flowers from ants provided perfect breeding grounds for mosquitos. Even in the sick wards themselves the legs of the beds were placed in shallow water, again to keep the ants away, and there were no screen in any of the windows or doors. Patients, furthermore, were placed in the wards according to nationality, rather than by disease, with the results that every ward had its malaria and yellow-fever cases. As Dr. Gorgas was to write, had the French been consciously trying to propagate malaria and yellow fever, they could not have provided conditions better suited for the purpose. (McCullough, 1977)

²History records that the French effort to build a canal across the isthmus did indeed fail, in part because of malaria.

³Williams (1951) presents a thorough history of the US Public Health Service.
(spraying, water management, screening, and quinine) and most of these demonstrations were highly successful, resulting in 70% declines in morbidity.

The federal government’s large-scale efforts against malaria in the South began with World War I. In previous wars, a significant portion of the troops were made unfit for service because of disease contracted on or around encampments. The PHS, working now with both a strong knowledge base on malaria control and greatly increased funding, undertook drainage and larviciding operations in Southern military camps as well as in surrounding areas. After the War, the IHB and PHS expanded the demonstration work further. By the mid-1920s, the boards of health of each state, following the IHB/PHS model, had taken up the mantle of the malaria control.

During this period, the South experienced a substantial decline in malaria. Malaria mortality per capita is seen in Panel A of Figure 1. Apart from a hiccup in the first years of the Depression, the region saw a drop of around 50% in the 15 years following WWI.

### 2.3 The Worldwide Campaign to Eradicate Malaria, circa 1950

While some of the innovations in malaria control diffused to less-developed regions, more technological advance was required before the poorer countries of the Americas were able to launch serious campaigns against malaria. These campaigns had a peculiar starting point: in 1941 Swiss chemists seeking to build a better mothball invented a chemical known today as DDT (short for dichloro-diphenyl-trichloro-ethane). Early tests showed this new chemical to be of extraordinary value as a pesticide: it rapidly killed a variety of insects and had no immediately apparent effects on mammals. DDT proved enormously valuable to the Allied war and occupation efforts in combatting typhus (transmitted by lice) and later malaria. The United Nations Reconstruction and Relief Agency used DDT in the late 1940s to essentially eradicate malaria from Sardinia, Italy, in the lapse of a few years.

The World Health Organization (WHO) proposed a worldwide campaign to eradicate malaria in the late 1940s and early 1950s. While the WHO mostly provided technical assistance and

---

3The historical narrative on the worldwide campaign is drawn from Harrison (1978).
moral suasion, substantial funding came from the USAID and UNICEF. The nations of Latin America took up this task in the 1950s. While individuals nations had formal control of the design and implementation of the programs, their activities were comparatively homogeneous as per the dictates of their international funders. The central component of these programs was the spraying of DDT, principally in the eaves of houses. Its purpose was not to kill every mosquito in the land, but rather to interrupt the transmission of malaria for long enough that the existing stock of parasites would die out. After that, the campaigns would go into a “maintenance” phase in which imported cases of malaria were to be managed medically.

The Latin American countries analyzed in the present study (Mexico, Colombia, and Brazil) all mounted malaria eradication campaigns, and all saw large declines in malaria prevalence. Panel B of Figure 1 shows malaria cases per capita in Colombia. A decline of approximately 80% is evident in the graph. The campaign ultimately proved inadequate to the task, and, in many areas, malaria partially resurfaced two decades later. But in almost all parts of the hemisphere, malaria never returned to its levels from before the application of DDT.

2.4 Research Design

The first factor in the research design is that the commencement of eradication was substantially due to factors external to the affected regions. The eradication campaign relied heavily upon critical innovations to knowledge from outside the affected areas. Such innovations were not related to or somehow in anticipation of the future growth prospects of the affected areas, and therefore should not be thought of as endogenous in this context. This contrasts with explanations that might have potentially troublesome endogeneity problems, such as, for example, positive income shocks in the endemic regions.

Second, the anti-malaria campaigns achieved considerable progress against the disease in less than a decade. This is a sudden change on historical time scales, especially when compared to trend changes in mortality throughout recent history, or relative to the gradual recession of malaria in the Midwestern US or Northern Europe. Moreover, I examine outcomes over a time span of 60 to 150 years of birth, which is unquestionably long relative to the malaria eradication campaigns.
The final element in the identification strategy is that different areas within each country had distinct incidences of malaria. In general terms, this meant that the residents of the US South, southern Mexico, northern Brazil, and lowland Colombia were relatively vulnerable to infection. Populations in areas with high (pre-existing) infection rates were in a position to benefit from the new treatments, whereas areas with low endemicity were not. This cross-regional difference permits a treatment/control strategy.

The advent of the eradication effort combines with the cross-area differences in pre-treatment malaria rates to form the research design. The variable of interest is the pre-eradication malaria intensity. By comparing the cross-cohort evolution of outcomes (e.g., adult income) across areas with distinct infection rates, one can assess the contribution of the eradication campaigns to the observed changes. (Specific estimating equations are presented below.)

How realistic is the assumption that areas with high infection rates benefited more from the eradication campaign? Mortality and morbidity data indicate drops of fifty to eighty percent in the decade following the advent of the eradication efforts. (See Figure 1.) Such a dramatic drop in the region’s average infection rate, barring a drastic reversal in the pattern of malaria incidence across the region, would have had the supposed effect of reducing infection rates more in highly infected areas than in areas with moderate infection rates. The decline in malaria incidence as a function of intensity prior to the eradication campaign is found in Figure 2. The basic assumption of the present study — that areas where malaria was highly endemic saw a greater drop in infection than areas with low infection rates — is borne out across areas in the countries where data are available.

---

4 Humid areas with slow-moving water were the preferred nursery for mosquitoes, the vector that transmits malaria.

5 This figure embodies the first-stage relationship. Consider the aggregate first-stage equation:

\[ M_{jt} = \gamma M_{j}^{pre} \times \text{Post}_t + \delta_j + \delta_t + \eta_{jt} \]

For area \( j \) in year \( t \). This equation can be written in first-differenced form and evaluated in the post-campaign period:

\[ \Delta M_{j}^{post} = \gamma M_{j}^{pre} + \text{constant} + \nu_{jt}, \]

an equation that relates the observable variables graphed in Figure 2.
3 Data Sources and Definitions

The micro-level data employed in the present study come from the Integrated Public Use Micro Sample (IPUMS), a project to harmonize the coding of census microdata from the United States and several other countries (Ruggles and Sobek (1997); Sobek et al. (2002)). I analyze the census data from the United States, Brazil, Colombia and Mexico.

The geographic units employed in this analysis are place of birth rather than current residence. Matching individuals with malaria rates of the area where they end up as adults would then be difficult to interpret because of selective migration. Instead, I use the information on malaria intensity in an individual’s area of birth to conduct the analysis. For the U.S., Mexico and Brazil, this means the state of birth. The Colombian census also contains information on birthplace by municipio, a second-order administrative unit similar to U.S. counties.

For the United States, the base sample consists of native-born white males in the Integrated Public Use Micro Sample or IPUMS (Ruggles and Sobek, 1997) and North Atlantic Population Project (NAPP, 2004) datasets between the ages of 25 and 55, inclusive, for the census years 1880-1990, which includes cohorts years of birth ranging from 1825 to 1965. I use two proxies for labor productivity that are available for a large number of Censuses. The occupational income score and Duncan socioeconomic index are both average indicators by disaggregated occupational categories that were calibrated using data from the 1950 Census. The former variable is the average by occupation of all reported labor earnings. The measure due to Duncan (1961) is instead a weighted average of earnings and education among males within each occupation. Both variables can therefore measure shifts in income that take place between occupations. The Duncan measure has the added benefit of picking up between-occupation shifts in skill requirements for jobs. Occupation has been measured by the Census for more than a century, and so these income proxies are available for a substantial stretch of cohorts.

The data on native-born males from the Brazilian and Mexican IPUMS-coded censuses from 1960 to 2000 are similarly pooled, resulting in birth cohorts from 1905 to 1975. These censuses contain questions on literacy, years of education and income. I also construct an income score based
on occupation and industry to better compare with the US results.

For Colombia, I use the IPUMS microdata on native males from the censuses of 1973 and 1993 (those for which municipio of birth was available). This yields birth cohorts from 1918 to 1968. I use the census-defined variables for literacy and years of schooling. I also use the income score defined from the Mexican and Brazilian data.

I combine microdata from various censuses to construct panels of average outcomes by cohort. Cohorts are defined by both when they were born and where they were born. To construct these panels, I pool the micro-level census data. The individual-level outcomes in the microdata are projected on to dummies for year-of-birth × census year × country. (Cohorts can appear in multiple censuses in this pooling strategy.) I then take the average residual from this procedure for each cell defined by period of birth and state (or municipio in the case of Colombia) of birth. For Section 5, I compare cohorts born well before or just after the campaign, so the period of birth is defined by childhood exposure to the eradication campaigns. In Section 4, I consider how cross-area outcomes changes by year of birth, so the panels are constructed with year of birth × area of birth as the units of observation.

Malaria data are drawn from a variety of sources. United States data are reported from by the Census (1894), Maxcy (1924) and later in the Vital Statistics (Census, 1933). Mexican data are drawn from Pesqueira (1957) and from the Mexican Anuario Estadístico (Dirección General de Estadística, 1960). SEM (1957) and the Colombian Anuario de Salubridad (DANE, 1970) are the sources for the Colombian data. Data on malaria ecology are derived from Gallup, Sachs and Mellinger (1999) and Poveda et al. (2000). The ecology data were matched with states and municipios using GIS.

4 Cohort-Specific Results

The shift in the malaria-income relationship coincides with childhood exposure to the eradication efforts. This can be seen graphically in this section. For each year of birth, OLS regression coefficients are estimated on the resulting cross section of states/municipios of birth. Consider a
simple regression model of an average outcome, \( Y_{jk} \), for a cohort with state of birth \( j \) and year of birth \( k \):

\[
Y_{jk} = \beta_k M_j + \delta_k + X_j \Gamma_k + \nu_{jk}
\]  

(1)
in which \( \beta_k \) is year-of-birth-specific coefficient on malaria, \( X_j \) is a vector of other state-of-birth controls,\(^6\) and \( \delta_k \) and \( \Gamma_k \) are cohort-specific intercept and slope coefficients. I estimate this equation using OLS for each year of birth \( k \). This specification allows us to examine how the relationship between income and pre-eradication malaria (\( \hat{\beta}_k \)) differs across cohorts.

I start with a simple graphical analysis using this flexible specification for cross-cohort comparison. Figures 4, 5, 6, and 7 display plots of the estimated \( \beta_k \), for the various outcomes and countries under study. The \( x \) axis is the cohort’s year of birth. The \( y \) axis for each graphic plots the estimated cohort-specific coefficients on the area-of-birth measure of malaria. Each cohort’s point estimate is marked with a dot.

Results for the US are shown in Figure 4, which displays the coefficient on state-of-birth 1890 malaria mortality for each year of birth. The additional variables in the summarized regressions include controls for health conditions and educational resources. (Appendix C has details on these variables. Section 5 below considers the sensitivity of these results to the choice of control sets.)

To consider the effects of childhood exposure to malaria, observe that US cohorts that were already adults in 1920 were too old to have benefited from the eradication efforts during childhood. On the other hand, later cohorts experienced reduced malaria infection during their childhood. This benefit increased with younger cohorts who were exposed to the anti-malaria efforts for a greater fraction of their childhood. The dashed lines therefore measure the number of years of potential childhood exposure\(^7\) to the malaria-eradication campaign. (The line is rescaled such that pre-1890 and post-1940 levels match those of the \( \hat{\beta}_k \). The exposure line is not rescaled in the \( x \) dimension.)

\(^6\)These additional controls are used in constructing the ultimate panels of Tables 3, 4, and 5.

\(^7\)Specifically, the formula is \( \text{Exp}_k = \max(\min(18, k - (1920 - 18)), 0) \), which treats 1920 as an approximate start date for exposure. Because the campaigns had their effect over a decade or more, the childhood-exposure measure represents an optimistically fast guess.
eradication campaign.

In the Latin American samples, the malaria-related change in outcomes across cohorts coincides roughly to childhood exposure to the campaigns, rather than a pre-existing trend. Figures 5, 6 and 7 display these results for Mexican and Brazilian states, and Colombian municipios, respectively. In each case, a trend break is evident approximately for those cohorts who were born just late enough to be exposed to the eradication campaign during some of their childhood.

Formal statistical tests confirm that the shift in the income/malaria ecology relationship coincided with exposure to malaria eradication, rather than with some trend or autoregressive process. This can be seen by treating the estimated $\beta_k$ as a time series and estimating the following regression equation:

$$\hat{\beta}_k = \alpha \text{Exp}_k + \sum_{i=1}^{n} \gamma_n k^n + \Phi(L)\beta_k + \eta_k + \text{constant} + \epsilon^t_t$$  \hspace{1cm} (2)

in which $\text{Exp}_k$ is exposure to malaria eradication (defined above), the $k^n$ terms are $n$th-order trends, and $\Phi(L)$ is a distributed lag operator. To account for the changing precision with which the generated observations are estimated, observations are weighted by the inverse of the standard error for $\hat{\beta}_k$. Table 1 reports estimates of equation 2 under a variety of order assumptions about trends and autoregression. The dependent variables are the cohort-specific regression estimates of outcomes on malaria that are shown in the figure above.

Panel A contains estimates for the United States. For the occupational income score, the estimates on the exposure term are broadly similar across specifications, and there is no statistically significant evidence of trends or autoregression in these $\beta_k$. When the Duncan SEI is used instead, there is evidence of a downward trend, but estimates of the exposure coefficient are stable once this is accounted for.

These point estimates imply reasonable reduced-form magnitudes for the effect of childhood exposure to malaria. States at the 90th and 10th percentiles of 1890 malaria mortality differed along this measure by 0.07% of total mortality. (The 10th-percentile states were essentially malaria free, while the 90th-percentile states had seven percent of their deaths attributed to malaria.) On the other hand, white males born in the South between 1875 and 1895 has average occupational
income and Duncan indices of 21 and 26, respectively. Therefore, these point estimates suggest
an reduced-form effect of around ten percent when comparing the 90th and 10th percentile states.
(These terms are reported in curly brackets in Table 1.)

In the Latin American samples, the malaria-related change in outcomes across cohorts coincides
roughly to childhood exposure to the campaigns, rather than a pre-existing trend. Figures 5, 7,
and 6 display these results for Brazilian states, Colombian municipios, and Mexican states, respec-
tively. In each case, trend breaks are visible approximately for those cohorts who were born just
late enough to be exposed to the eradication campaign during some of their childhood.

[MORE TO COME]

5 Pre/Post Comparisons

I compare changes in socioeconomic outcomes by cohort across areas with distinct malaria intensi-
ties, in order to assess the contribution of the eradication campaign to the observed changes. The
basic equation to be estimated is

$$\Delta Y_{i,t} = \beta M_{i,t-1} + X_{i,t-1}\Gamma + \alpha + \varepsilon_{i,t}$$

in which \(Y\) is some socioeconomic outcome for state or municipio \(i\). The time subscript \(t\) refers to
a year of birth following the malaria-eradication campaign, while \(t - 1\) indicates being born (and
having become an adult) prior to advent of the campaign. The pre-program malaria incidence is
\(M_{i,t-1}\), the \(X\) variables are a series of controls, and \(\alpha\) is a constant term. The parameter of interest
is \(\beta\). This parameter can be thought of as coming from a reduced-form equation, in the sense of
two-stage least squares.8

Areas in the US with higher malaria burdens prior to the eradication efforts saw larger cross-

8The model is derived as follows. Consider an individual \(i\), in area \(j\), with year-of-birth \(t\), we start with an
individual-level model with individual infection data and linear effects of malaria:

$$Y_{ijt} = \alpha M_{ijt} + \delta_j + \delta_t + \varepsilon_{ijt}$$

where \(M_{ijt}\) is a measure of childhood malaria infection. No data set has both childhood malaria infection data and
adult income, and the research design is fundamentally at the period-of-birth \(\times\) area-of-birth level, so I rewrite the
cohort growth rates in income, as measured by the occupational proxies. These results are found in Table 3. Panel A contain estimates for the basic specification of equation 5, plus a dummy for being born in the South. If the oldest cohorts had high malaria infection and low productivity because of some mean-reverting shock, we might expect income gains for the subsequent cohorts even in the absence of a direct effect of malaria on productivity. I control for the natural logarithm of state wages by using data on labor earnings by state in 1899 from Lebergott (1964). Panel B contains the basic mean-reversion control, while Panel C includes a more flexible control for wages. Panel D controls for additional measures of health, while Panel E includes controls for fraction urban and black, and for the 1930 unemployment rate. Panel F contains controls for changes in educational policy and pre-existing literacy rates, while Panel G includes all of the above control variables simultaneously in the specification. The estimates for malaria are not substantially affected by the inclusion of these additional variables. Figure 8 displays a scatter plot of the orthogonal component of cross-cohort income growth versus malaria, after having projecting each variable onto the control variables.

Table 4 reports the estimates for Mexico and Brazil. Malarious areas saw faster cross-cohort growth in income and literacy, but mixed evidence on years of schooling. Results are shown for a variety of control variables, including sectoral mix, infant mortality, and proxies for economic development. Figures 9 and 11 displays the orthogonal component of malaria and changes across cohort in these outcomes.

Results from Colombia indicate that childhood exposure to malaria suppressed income. Cross-cohort growth in income, literacy and education was higher in the areas with more perverse malaria ecology, as shown in Table 5. These estimates are robust to including a variety of controls for sectoral

equation above in aggregate form:
\[ Y_{jt} = \tilde{\alpha}M_{jt} + \delta_j + \delta_t + \varepsilon_{jt} \]

I partition the cohorts into those born after the advent of the campaign and those who were already adults by the time the campaign started. I then difference the model along these lines, and take \( M_{i,t-1} \) as an instrument for the decline in malaria. The resulting reduced form of this system is equation 5. Alternatively, one could have written the individual-level model with separate terms for individual and aggregate infection variables, the latter of which reflecting some spillover from peer infection to own human capital. But both of these effects would be subsumed into the \( \tilde{\alpha} \) coefficient on the ecological infection rate, and it is this composite coefficient that I seek to measure in the present study.
mix, violence, and proxies of economic development. Comparisons of effects across 90/10 percentile differences in malaria are broadly similar, especially when temperature and altitude are used as instruments to correct for measurement error, which is likely large especially in the measure of cases notified. The residualized components of the cross-cohort changes and malaria ecology (using the Poveda measure) are shown in Figure 10.

6 Conclusions

This study considers the socioeconomic impact of the malaria-eradication campaigns in the United States (circa 1920), and in Brazil, Colombia and Mexico (circa 1955). The goal is to measure how much childhood exposure to malaria depresses labor productivity.

Several factors combine to form the research design. The eradication campaigns studied happened because of advances in medical and public-health knowledge, which mitigates concerns about reverse causality of the timing of eradication efforts. Highly malarious areas saw large drops in their malaria incidence following the campaign. Furthermore, these gains against the disease were realized in approximately a decade. Data from regional malaria eradication programs are collected and collated with publicly available census data.

In both absolute terms and relative to the comparison group of non-malarious areas, cohorts born after eradication had higher income as adults than the preceding generation. Similar increases in literacy and the returns to schooling are observed. Mixed results are found for years of schooling, consistent with the economic theory of schooling (which compares returns with opportunity costs).

References


——— (2000). División Político-Administrativa de Colombia, Bogotá, Colombia: DANE.


Figure 1: Malaria Incidence Before and After the Eradication Campaigns

Panel A: Mortality per Capita, Southern United States

Panel B: Cases Notified per Capita, Colombia

Notes: Panel A plots the estimated malaria mortality per capita for the Southern region and bordering states. Because the death registration system was being phased in over the period, a regression model with state fixed effects is used to control for sample changes. Panel B reports data on notified cases of malaria for Colombia (SEM, 1979).
Figure 2: Highly Infected Areas Saw Greater Declines in Malaria

Notes: The y axis displays the estimated decrease in malaria mortality post-intervention. The x axis is the pre-campaign malaria mortality rate. The 45-degree line represents complete eradication. Both variables are expressed per 100,000 population. United States data are reported in Maxcy (1924) and Vital Statistics (Census, 1933). Mexican data are drawn from Pesqueira (1957) and from the Mexican Anuario Estadistico (Dirección General de Estadística, 1960). SEM (1957) and the Colombian Anuario de Salubridad (DANE, 1968-70) are the sources for the Colombian data.
Figure 3: Childhood Exposure to Eradication Campaign

Notes: This graph displays the fraction of childhood that is exposed to a hypothetical (and nearly instantaneous) campaign as a function of year of birth minus the start year of the campaign.
Notes: These graphics summarize regressions of income proxies on pre-eradication malaria-mortality rates (measured by the Census in 1890). The $y$ axis for each graphic plots the estimated cohort-specific coefficients on the state-level malaria measure. The $x$ axis is the cohort’s year of birth. Each cohort’s point estimate is marked with a dot. The dashed lines measure the number of years of potential childhood exposure to the malaria-eradication activities in the South. For each year-of-birth cohort, OLS regressions coefficients are estimated on the cross section of states of birth. The state-of-birth average outcome is regressed on to malaria, Lebergott’s measure of 1899 wage levels, a dummy for the Southern region and the various control variables described in Appendix C.
Figure 5: Cohort-Specific Relationship: States in Brazil

Literacy

Years of Schooling

Total Income

Earned Income

Notes: These graphics summarize regressions of the indicated socioeconomic outcomes on malaria-ecology rates (measured by Mellinger et al. (1998)). The y axis for each graphic plots the estimated cohort-specific coefficients on the state-level malaria measure. The x axis is the cohort’s year of birth. Each cohort’s point estimate is marked with a dot. The dashed lines measure the number of years of potential childhood exposure to the commencement of the malaria-eradication campaign in Latin America. For each year-of-birth cohort, OLS regressions coefficients are estimated on the cross section of states of birth. The state-of-birth average outcome is regressed on to malaria, region dummies, and the various control variables described in Appendix C.
Figure 6: Cohort-Specific Relationship: Municipios in Colombia

Notes: These graphics summarize regressions of the indicated socioeconomic outcomes on malaria-ecology rates (measured by Poveda et al. (2000)). The y axis for each graphic plots the estimated cohort-specific coefficients on the municipio-level malaria measure. The x axis is the cohort’s year of birth. Each cohort’s point estimate is marked with a dot. The dashed lines measure the number of years of potential childhood exposure to the commencement of the malaria-eradication campaign in Latin America. For each year-of-birth cohort, OLS regressions coefficients are estimated on the cross section of municipios of birth. The state-of-birth average outcome is regressed on to malaria, region dummies, and the various control variables described in Appendix C.
Figure 7: Cohort-Specific Relationship: States in Mexico

Notes: These graphics summarize regressions of the indicated socioeconomic outcomes on malaria-mortality rates (measured by Pesquiera (1957)). The y axis for each graphic plots the estimated cohort-specific coefficients on the state-level malaria measure. The x axis is the cohort's year of birth. Each cohort's point estimate is marked with a dot. The dashed lines measure the number of years of potential childhood exposure to the commencement of the malaria-eradication campaign in Latin America. For each year-of-birth cohort, OLS regressions coefficients are estimated on the cross section of states of birth. The state-of-birth average outcome is regressed on to malaria, region dummies, and the various control variables described in Appendix C.
Figure 8: Cross-Cohort Growth Rates versus Malaria: US States

Panel A: Change in Occupational Income Score

Panel B: Change in Duncan Socio-Economic Indicator

Notes: Top panel displays results for the occupational income score, while the bottom panel uses the Duncan Socioeconomic Indicator. The y-axis are the changes in the indicated income proxy between cohorts born before 1895 and those born after 1925. The x-axis plots malaria mortality over total deaths in 1890. Both variables are residuals from having projected the original data on to a dummy for South, a 4th-order polynomial for Lebergott 1899 wage series, child-mortality rate in 1890, urbanization in 1910, adult literacy in 1910, doctors per capita in 1898, state public health spending in 1898, hookworm infection circa 1917, fraction black in 1910, unemployment rate in 1930, and the log change from 1905-25 in school-term length, pupil/teacher ratio, and teacher salary.
Figure 9: Cross-Cohort Growth Rates versus Malaria: States in Brazil

Notes: The y-axis are the changes in the indicated socioeconomic variable between cohorts born before 1940 and those born after 1957. The x-axis plots Mellinger measure of malaria ecology. Both variables are residuals from having projected the original data on to the various control variables described in Appendix C.
Figure 10: Cross-Cohort Growth Rates versus Malaria: Municipios in Colombia

Notes: The y-axis are the changes in the indicated socioeconomic variable between cohorts born before 1940 and those born after 1957. The x-axis plots Poveda measure of malaria ecology. Both variables are residuals from having projected the original data on to the various control variables described in Appendix C.
Notes: The $y$-axis are the changes in the indicated socioeconomic variable between cohorts born before 1940 and those born after 1957. The $x$ axis measures malaria-mortality rates (measured by Pesquiera (1957)). Both variables are residuals from having projected the original data on to the various control variables described in Appendix C.
### Table 1: Exposure to Malaria Eradication versus Alternative Time-Series Processes

<table>
<thead>
<tr>
<th>Outcome Variables:</th>
<th>Time-Series Estimates of the Exposure Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel A: United States</td>
</tr>
<tr>
<td>Occupational Income Score</td>
<td>28.684 * 33.802 * 18.470 * 34.611 * 14.554 * 18.783 *</td>
</tr>
<tr>
<td></td>
<td>(1.509) (3.664) (2.700) (4.105) (3.196) (5.286)</td>
</tr>
<tr>
<td></td>
<td>{0.095} {0.112} {0.061} {0.115} {0.048} {0.062}</td>
</tr>
<tr>
<td>Duncan’s Socioeconomic Index</td>
<td>52.549 * 48.862 * 32.486 * 57.078 * 27.900 * 32.744 *</td>
</tr>
<tr>
<td></td>
<td>{0.138} {0.085} {0.105} {0.073} {0.086}</td>
</tr>
<tr>
<td></td>
<td>Panel B: Brazil</td>
</tr>
<tr>
<td></td>
<td>Literacy</td>
</tr>
<tr>
<td></td>
<td>0.029 * 0.018 * 0.021 * 0.017 * 0.018 * 0.012 *</td>
</tr>
<tr>
<td></td>
<td>(0.002) (0.004) (0.003) (0.004) (0.004) (0.005)</td>
</tr>
<tr>
<td></td>
<td>{0.141} {0.087} {0.101} {0.085} {0.089} {0.061}</td>
</tr>
<tr>
<td></td>
<td>Years of Schooling</td>
</tr>
<tr>
<td></td>
<td>0.214 * 0.116 0.162 * 0.349 * 0.068 0.252 ~</td>
</tr>
<tr>
<td></td>
<td>(0.025) (0.070) (0.047) (0.057) (0.041) (0.099)</td>
</tr>
<tr>
<td></td>
<td>{1.047} {0.567} {0.792} {1.707} {0.333} {1.233}</td>
</tr>
<tr>
<td></td>
<td>Log Total Income</td>
</tr>
<tr>
<td></td>
<td>0.073 * 0.094 * 0.061 * 0.104 * 0.054 * 0.087 *</td>
</tr>
<tr>
<td></td>
<td>(0.005) (0.011) (0.011) (0.011) (0.014) (0.019)</td>
</tr>
<tr>
<td></td>
<td>{0.357} {0.460} {0.298} {0.509} {0.264} {0.426}</td>
</tr>
<tr>
<td></td>
<td>Log Earned Income</td>
</tr>
<tr>
<td></td>
<td>0.056 * 0.080 * 0.061 * 0.082 * 0.058 * 0.076 ~</td>
</tr>
<tr>
<td></td>
<td>(0.008) (0.022) (0.014) (0.025) (0.016) (0.037)</td>
</tr>
<tr>
<td></td>
<td>{0.274} {0.391} {0.298} {0.401} {0.284} {0.372}</td>
</tr>
<tr>
<td></td>
<td>Panel C: Colombia</td>
</tr>
<tr>
<td></td>
<td>Literacy</td>
</tr>
<tr>
<td></td>
<td>0.099 * 0.048 0.082 * 0.067 ~ 0.055 * 0.048 *</td>
</tr>
<tr>
<td></td>
<td>(0.010) (0.027) (0.016) (0.028) (0.016) (0.027)</td>
</tr>
<tr>
<td></td>
<td>{0.039} {0.019} {0.032} {0.027} {0.022} {0.019}</td>
</tr>
<tr>
<td></td>
<td>Years of Schooling</td>
</tr>
<tr>
<td></td>
<td>0.736 * 1.537 0.602 * 0.550 ~ 0.458 * 0.661 ~</td>
</tr>
<tr>
<td></td>
<td>(0.100) (0.225) (0.148) (0.268) (0.165) (0.311)</td>
</tr>
<tr>
<td></td>
<td>{0.292} {0.457} {0.239} {0.218} {0.181} {0.262}</td>
</tr>
<tr>
<td></td>
<td>Industrial Income Score</td>
</tr>
<tr>
<td></td>
<td>0.143 * 0.090 ~ 0.096 ~ 0.105 ~ 0.070 ~ 0.108 ~</td>
</tr>
<tr>
<td></td>
<td>(0.015) (0.041) (0.020) (0.035) (0.021) (0.040)</td>
</tr>
<tr>
<td></td>
<td>{0.057} {0.036} {0.038} {0.042} {0.028} {0.043}</td>
</tr>
<tr>
<td></td>
<td>Panel D: Mexico</td>
</tr>
<tr>
<td></td>
<td>Literacy</td>
</tr>
<tr>
<td></td>
<td>0.008 * -0.006 0.009 * -0.009 0.006 ~ -0.007</td>
</tr>
<tr>
<td></td>
<td>(0.003) (0.004) (0.003) (0.005) (0.003) (0.004)</td>
</tr>
<tr>
<td></td>
<td>{0.026} {-0.020} {0.026} {-0.027} {0.018} {-0.022}</td>
</tr>
<tr>
<td></td>
<td>Years of Schooling</td>
</tr>
<tr>
<td></td>
<td>-0.087 * -0.194 * -0.086 * -0.178 * -0.087 * -0.172 *</td>
</tr>
<tr>
<td></td>
<td>(0.020) (0.051) (0.024) (0.046) (0.026) (0.066)</td>
</tr>
<tr>
<td></td>
<td>{-0.267} {-0.594} {-0.264} {-0.545} {-0.267} {-0.527}</td>
</tr>
<tr>
<td></td>
<td>Log Earned Income</td>
</tr>
<tr>
<td></td>
<td>0.067 * 0.021 0.085 * 0.036 0.064 ~ 0.027</td>
</tr>
<tr>
<td></td>
<td>(0.016) (0.035) (0.017) (0.026) (0.020) (0.027)</td>
</tr>
<tr>
<td></td>
<td>{0.205} {0.064} {0.260} {0.110} {0.196} {0.083}</td>
</tr>
</tbody>
</table>

**Regression Specifications:**

- **Order of Polynomial Trend:** 0 1 0 2 0 2
- **Order of Autoregressive Process:** 0 0 1 0 2 2

**Notes:** This table reports estimates of equation 2 using OLS. The outcome variables used to construct the time series of \( \hat{\beta}_k \) are as indicated in each row. Robust (Huber-White) standard errors in parentheses. Single asterisk denotes statistical significance at the 99% level of confidence; tilde at the 95% level. Observations are weighted according the inverse of the coefficient’s standard error. Reporting of additional terms suppressed. The terms in curly brackets report the point estimate multiplied by the difference between 90th and 10th percentile malaria intensity. For the United States, this number is also normalized by the average value of the relevant income proxy for white males born in the South between 1875 and 1895.
<table>
<thead>
<tr>
<th>Income Variable:</th>
<th>Time-Series Estimates of the Exposure Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel A: Brazil</td>
</tr>
<tr>
<td>Log Total Income</td>
<td>-0.359 *** 1.326 *** -0.122 0.712 *** -0.072 0.543 ***</td>
</tr>
<tr>
<td></td>
<td>(0.102) (0.339) (0.077) (0.185) (0.072) (0.193)</td>
</tr>
<tr>
<td></td>
<td>{-0.018} [0.065] [-0.006] [0.035] [-0.004] [0.027]</td>
</tr>
<tr>
<td>Log Earned Income</td>
<td>0.145 * 0.224 0.168 * 0.265 0.196 ** 0.156</td>
</tr>
<tr>
<td></td>
<td>(0.087) (0.229) (0.089) (0.246) (0.088) (0.250)</td>
</tr>
<tr>
<td></td>
<td>{0.007} [0.011] [0.008] [0.013] [0.010] [0.008]</td>
</tr>
<tr>
<td></td>
<td>Panel B: Mexico</td>
</tr>
<tr>
<td>Log Earned Income</td>
<td>-0.247 0.884 -0.281 0.566 -0.223 0.716</td>
</tr>
<tr>
<td></td>
<td>(0.432) (0.956) (0.440) (0.750) (0.433) (0.843)</td>
</tr>
<tr>
<td></td>
<td>{-0.008} [0.027] [-0.009] [0.017] [-0.007] [0.022]</td>
</tr>
</tbody>
</table>

Regression Specifications:

<table>
<thead>
<tr>
<th>Order of Polynomial Trend:</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>0</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order of Autoregressive Process:</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of equation 2 using OLS. The dependent variables are time series of cohort-specific interactions between of the malaria variable with years of schooling. Robust (Huber-White) standard errors in parentheses. Single asterisk denotes statistical significance at the 99% level of confidence; tilde at the 95% level. Observations are weighted according the inverse of the coefficient’s standard error. The outcome variables used to construct the time series of $\hat{\beta}_k$ are as indicated in each row. The underlying regressions are run separately on the micro data for each year of birth, and include a full set of dummies for years of schooling and state of birth and linear interactions of years of schooling with region of birth and the control variables in the “full controls” specifications of Table 4. The terms in curly brackets report the point estimate multiplied by the difference between 90th and 10th percentile malaria intensity. Reporting of additional terms suppressed.
Table 3: Cross-Cohort Growth versus Malaria: United States

<table>
<thead>
<tr>
<th>Additional Controls</th>
<th>Occupational Income Score</th>
<th>Duncan’s SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>37.686 ***</td>
<td>60.899 ***</td>
</tr>
<tr>
<td></td>
<td>(11.036)</td>
<td>(21.476)</td>
</tr>
<tr>
<td></td>
<td>{0.125}</td>
<td>{0.160}</td>
</tr>
<tr>
<td>Wages, 1899 (Lebergott)</td>
<td>37.927 ***</td>
<td>60.316 ***</td>
</tr>
<tr>
<td></td>
<td>(11.101)</td>
<td>(21.311)</td>
</tr>
<tr>
<td></td>
<td>{0.126}</td>
<td>{0.158}</td>
</tr>
<tr>
<td>Wages as 4th order polynomial</td>
<td>36.617 ***</td>
<td>55.824 ***</td>
</tr>
<tr>
<td></td>
<td>(10.763)</td>
<td>(19.909)</td>
</tr>
<tr>
<td></td>
<td>{0.121}</td>
<td>{0.147}</td>
</tr>
<tr>
<td>Health</td>
<td>33.897 ***</td>
<td>63.480 ***</td>
</tr>
<tr>
<td></td>
<td>(9.733)</td>
<td>(20.610)</td>
</tr>
<tr>
<td></td>
<td>{0.112}</td>
<td>{0.167}</td>
</tr>
<tr>
<td>Education</td>
<td>44.825 ***</td>
<td>59.306 **</td>
</tr>
<tr>
<td></td>
<td>(12.240)</td>
<td>(23.279)</td>
</tr>
<tr>
<td></td>
<td>{0.148}</td>
<td>{0.156}</td>
</tr>
<tr>
<td>Other</td>
<td>30.118 ***</td>
<td>45.827 **</td>
</tr>
<tr>
<td></td>
<td>(11.400)</td>
<td>(18.134)</td>
</tr>
<tr>
<td></td>
<td>{0.100}</td>
<td>{0.120}</td>
</tr>
<tr>
<td>Full controls</td>
<td>33.392 **</td>
<td>59.257 **</td>
</tr>
<tr>
<td></td>
<td>(13.844)</td>
<td>(29.103)</td>
</tr>
<tr>
<td></td>
<td>{0.111}</td>
<td>{0.156}</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of equation 5 using OLS. The units of observation are US states. The dependent variables are as indicated in the column headings. Robust (Huber-White) standard errors in parentheses. Single asterisk denotes statistical significance at the 99% level of confidence; tilde at the 95% level. Reporting of constant term suppressed. Unexposed cohorts are those born before 1890 and fully exposed cohorts are those born after 1920. Cohorts are determined based on state of birth. The universe for the base sample consists of the native-born white population between the ages of 25 and 55 (15–55 for literacy) in the 1880–2000 census microdata from the IPUMS and NAPP databases. The terms in curly brackets report the point estimate multiplied by the difference between 90th and 10th percentile malaria intensity and normalized by the average value of the relevant income proxy for white males born in the South between 1875 and 1895.
### Table 4: Cross-Cohort Growth versus Malaria: Mexico and Brazil

<table>
<thead>
<tr>
<th>Specification</th>
<th>Brazilian States (N=24)</th>
<th>Mexican States (N=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Literacy</td>
<td>Education</td>
</tr>
<tr>
<td>Full Controls</td>
<td>0.028 ***</td>
<td>0.190 **</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.093)</td>
</tr>
<tr>
<td></td>
<td>[0.137]</td>
<td>[0.929]</td>
</tr>
<tr>
<td>Drop Infant Mortality</td>
<td>0.026 ***</td>
<td>0.244 **</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.120)</td>
</tr>
<tr>
<td></td>
<td>[0.127]</td>
<td>[1.194]</td>
</tr>
<tr>
<td>Drop Sector</td>
<td>0.005</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.130)</td>
</tr>
<tr>
<td></td>
<td>[0.024]</td>
<td>[0.802]</td>
</tr>
<tr>
<td>Drop Mean Reversion</td>
<td>0.025 ***</td>
<td>0.259 **</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.103)</td>
</tr>
<tr>
<td></td>
<td>[0.122]</td>
<td>[1.267]</td>
</tr>
<tr>
<td>Drop All</td>
<td>0.022 ***</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.096)</td>
</tr>
<tr>
<td></td>
<td>[0.108]</td>
<td>[-0.171]</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of malaria in equation 5 using OLS. The units of observation are Mexican and Brazilian states. The dependent variables are as indicated in the column headings. Robust (Huber-White) standard errors in parentheses. Single asterisk denotes statistical significance at the 99% level of confidence; tilde at the 95% level. The terms in curly brackets report the point estimate multiplied by the difference between 90th and 10th percentile malaria intensity. Reporting of constant term suppressed. Unexposed cohorts are those born before 1940 and fully exposed cohorts are those born after 1960. Cohorts are determined based on state of birth. The universe for the base sample consists of the native-born population between the ages of 25 and 55 (15–55 for literacy) in the 1960–2000 census microdata from the IPUMS. The malaria measure for Brazil is Mellinger's ecology variable, while for Mexico it is malaria mortality circa 1950 (Pesqueira, 1957). All regressions include dummies for region of birth.
Table 5: Cross-Cohort Growth versus Malaria: Colombia

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Literacy</td>
<td>Years of</td>
<td>Income</td>
</tr>
<tr>
<td>OLS, Basic Specification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.089 ***</td>
<td>0.425 *</td>
<td>0.162 ***</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.222)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>2SLS, Temperature and Altitude Instruments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.093 **</td>
<td>0.940 ***</td>
<td>0.232 ***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.344)</td>
<td>(0.043)</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.372)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>Additional Controls:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conflict</td>
<td>0.081 ***</td>
<td>0.442 *</td>
<td>0.158 ***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.277)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Economic Activity</td>
<td>0.019</td>
<td>0.489 **</td>
<td>0.145 ***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.226)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Other Diseases</td>
<td>0.061 *</td>
<td>0.454 **</td>
<td>0.164 ***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.224)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Full Controls</td>
<td>0.014 ***</td>
<td>0.416 *</td>
<td>0.162 ***</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.240)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Instrumental Variables:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Other Two Proxies</td>
<td>0.317 ***</td>
<td>0.286</td>
<td>0.212 ***</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td>(0.480)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>Holdridge Climate Zone</td>
<td>0.113 **</td>
<td>0.764 *</td>
<td>0.258 ***</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.402)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>All of the Above Instruments</td>
<td>0.124 ***</td>
<td>0.816 ***</td>
<td>0.233 ***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.307)</td>
<td>(0.041)</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of malaria in equation 5 using OLS for the indicated dependent variables. The units of observation are Colombian municipios. The malaria variables are as indicated in the column headings. Robust (Huber-White) standard errors in parentheses. Single asterisk denotes statistical significance at the 99% level of confidence; tilde at the 95% level. To the present results on a similar scale, the terms in curly brackets reported in Panel A are the point estimates multiplied by the 90-10 difference across municipios in the malaria variable. Reporting of additional estimates is suppressed. Unexposed cohorts are those born before 1940 and fully exposed cohorts are those born after 1960. Cohorts are determined based on municipio of birth. The universe for the base sample consists of the native-born population between the ages of 25 and 55 (15–55 for literacy) in the 1973 and 1993 census microdata from the IPUMS. All regressions include dummies for region of birth.
A Construction of the Cohort-Level Data

The micro data for the RC analysis are drawn primarily from the IPUMS data for the United States, Brazil, Colombia, and Mexico. For each country, these data are used to construct a panel of income by year and area of birth. The cohort-level outcomes are constructed as follows.

1. The microdata are first pooled together.
2. The individual-level outcome variable are projected on to dummies for year-of-birth $\times$ Census year, i.e. I run the following regression:
   \[ y_{itk} = \delta_{tk} + \epsilon_{itk} \]
   for individual $i$ in cohort $k$ when observed in census year $t$. This regression absorbs all cohort, age, and period effects that are common for the whole country.
3. I then define cells for each combination of year of birth and area of birth. Within each cell, I compute the average of the estimated residuals (the $\hat{\epsilon}_{itk}$). Because these averages are constructed with differing degrees of precision, I also compute the square root of the cell sizes to use as weights when estimating equation 1.
4. I repeat this process separately for each outcome variable.

These average outcomes by cohort form the dependent variables in Section 4 and specifically Figures 4 through 7.

A.1 Details for the United States Sample

The underlying sample used for the United States consists of native-born white males in the age range $[25,60]$ in the 1900–1990 IPUMS microdata or in the 1880 microdata from the North Atlantic Population Project (NAPP, 2004). (These data were last accessed November 14, 2005.) This results in a data set with year-of-birth cohorts from 1825 to 1965. The original micro-level variables are defined as follows:

- **Occupational income score.** The occupational income score is an indicator of income by disaggregated occupational categories. It was calibrated using data from the 1950 Census, and is the average by occupation of all reported labor earnings. See Ruggles and Sobek (1997) for further details.

- **Duncan socio-economic index.** This measure is a weighted average of earnings and education among males within each occupation. The weights are based on analysis by Duncan (1961) who regressed a measure of perceived prestige of several occupations on its average income and education. This measure serves to proxy for both the income and skill requirements in each occupation. It was similarly calibrated using data from the 1950 Census.

For the majority of the years of birth, I can compute average income proxies for all of the 51 states plus the District of Columbia. The availability of state-level malaria data and the control variables restricts the sample further to 46 states of birth. Alaska, Colorado, the District of Colombia, Hawaii, and Oklahoma are excluded because of missing data for at least one of the other dependent variables. This leaves 46 states of birth in the base sample.

There are a number of cohorts born before 1885 for which as few as 37 states of birth are represented. (See Appendix Figure 1.) For those born between 1855 and 1885, this appears to be due to small samples, because, while the NAPP data are a 100% sample for 1880, there are no microdata for 1890 and 1900 IPUMS data are only a 1% sample. On the other hand, for the 1843-1855 birth cohorts, all but two of the years have all 46 states represented. Nevertheless, even with the 100% sample from 1880, there are as many as six states per year missing for those cohorts born before 1843. A number of the territories (all of which would later become states) were being first settled by people of European descent during the first half of the XIXth century, and it is quite possible that, in certain years, no one eligible to be enumerated was born in some territories. (Untaxed Indians were not counted in the censuses.) Note that I use the term state above to
refer to states or territories. Territories were valid areas of birth in the earlier censuses, and are coded in
the same way as if they had been states.

While this procedure generates an unbalanced panel, results are similar when using a balanced panel with
only those states of birth with the maximum of 141 valid observations. A comparison of the cohort-specific
estimates from the balanced and unbalanced panels shows high correlation (over 0.96, for example, in the
case of the full-controls specification for the occupational income score).

A.2 Details for the Brazilian Sample

The underlying sample used for Brazil consists of native males in the age range \([15,60]\) in the 1960–2000
IPUMS microdata. (These data were last accessed April 7, 2006.) This results in a data set with year-of-birth
cohorts from 1905 to 1984.

State is birth is available for these samples. Brazilian states (and several territories that were to become
states) were, by and large, consistently defined over the course of the sample. Those few that were not were
merged together to reflect administrative divisions in the early 1950s. Specifically, I merged Rondonia into
Guapure, Roraima into Rio Branco, Tocantins into Goias, Fernando de Noronha into Pernambuco, Serra do
Ainores into Minas Gerais, and Mato Grosso do Sul into Mato Grosso.

The original micro-level variables are as follows:

- **Literacy.** A binary variable individual measuring whether an individual can read and write at least
  a simple note.

- **Years of Schooling.** Numbers of years of education corresponding to highest grade completed.
  Non-numeric responses (e.g., “some secondary”) are mapped onto the midpoints of the appropriate
  intervals.

- **Total Income.** Records the total personal income from all sources in the prior month. In the
  empirical work above, this variable is treated in natural logs. This variable is intervalled in the 1960
  census, and their midpoints are used in translating the data into income.

- **Earned Income.** Records the personal income from their labor (wages, business, or farm) in the
  prior month. In the empirical work above, this variable is treated in natural logs.

A.3 Details for the Colombian Sample

The underlying sample used for Colombia consists of native males in the age range \([15,60]\) in the 1973 and
1991 IPUMS microdata. (These data were last accessed April 10, 2006.) This results in a data set with year-of-birth
cohorts from 1918 to 1977.

Area of birth is available in these samples at the level of departamento and municipio. The departamento
is a first-order administrative division, similar to a state, while the municipio is a second-order division,
similar to a county in the United States. A cohort’s municipio of birth is used in the present study to
construct a proxy for childhood exposure to malaria. Colombia contains over one thousand municipios in
the present day, but, to preserve confidentiality in the IPUMS data, some of the smaller municipios are
aggregated into larger groupings. This results in over 500 unique codes for area of birth, and I refer to these
units simply as “municipios” in the text. Because municipal boundaries change over time, maps (SEM, 1957)
and other administrative information (DANE, 2000) were used to relate data observed at various points in
time onto the IPUMS recode of municipio.

The original micro-level variables are as follows:

- **Literacy.** A binary variable individual measuring whether an individual can read and write.

- **Years of Schooling.** Numbers of years of education corresponding to highest grade completed.
  Non-numeric responses (e.g., “some secondary”) are mapped onto the midpoints of the appropriate
  intervals.
• **Industrial Income Score.** The industrial income score is an indicator of income by industry and class of worker. It was calibrated using data from the Brazilian and Mexican censuses for all available years. To remove census-year times country effects, the starting point for this variable is the residualized total income ($\hat{\epsilon}_{itk}$) described above, which is then averaged by industry and matched onto the Colombian sample. Because of the way this score is constructed, the variable is measured in natural logs. (Total income is available in the 1973 Colombian census, but the range of years of birth that these data cover is too limited.)

A.4 Details for the Mexican Sample

The underlying sample used for Mexico consists of native males in the age range [15,60] in the 1960–2000 IPUMS microdata. (These data were last accessed April 7, 2006.) This results in a data set with year-of-birth cohorts from 1905 to 1984.

State is birth is available for these samples. Mexican states (some of which were territories early on) were defined consistently throughout the sample period.

The original micro-level variables are as follows:

- **Literacy.** A binary variable individual measuring whether an individual can read and write.
- **Years of Schooling.** Numbers of years of education corresponding to highest grade completed. Non-numeric responses (e.g., “some secondary”) are mapped onto the midpoints of the appropriate intervals.
- **Earned Income.** Records the personal income from their labor (wages, business, or farm) in the prior month. In the empirical work above, this variable is treated in natural logs. (Total income is available in certain years of the Mexican censuses, but the range of years of birth that these data cover is inappropriate for the analysis.)
These graphs report additional summary statistics by year of birth for the \( \hat{\beta}_t \) reported in Figure 4 in the subplot labelled “Occupational Income Score; Full controls”.

These graphs report additional summary statistics by year of birth for the \( \hat{\beta}_t \) reported in Figure 5 in the subplot labelled “Total Income”.

38
Figure A – 3: Sample Statistics for the Colombian Sample

These graphs report additional summary statistics by year of birth for the $\hat{\beta}_t$ reported in Figure 6 in the subplot labelled “Income Score”.

Figure A – 4: Sample Statistics for the Mexican Sample

These graphs report additional summary statistics by year of birth for the $\hat{\beta}_t$ reported in Figure 7 in the subplot labelled “Earned Income”.

39
B Sources and Construction of the Malaria Data

Sources are indicated in parentheses at the end of each item.

- **United States.** Malaria mortality, expressed as a fraction of total mortality. This was measured in the 1890 Census as refers to the proceeding year. These data were collected by Census enumerators. (Bureau of the Census, 1894.)

- **Brazil.** An index of malaria ecology, computed using information on climate and local vectorial capacity. The construction of these data are described in Gallup, Sachs and, Mellinger (1999a). The source data were provided as raster data in one-degree grids. A GIS program was used to extract average malaria ecology by state. (Andrew Mellinger, private communication, and author’s calculations.)

- **Colombia.** Two measures of ecology are used, as well as one measure of morbidity. The Poveda measure is an index of malaria ecology based on climatic factors, described by Poveda, Graham and Epstein (2000). The authors would not provide the source data, so the map in that article displaying the computed survival probability of *p. vivax* (Fig 6.5) was digitized and fed into a GIS program, which was then used to construct averages by municipio. The Mellinger measure of malaria ecology is the same as that used for Brazil, and was averaged by municipio in a GIS program (the “Spatial Analyst” toolbox within ArcView). Glenn Hyman of the Centro Internacional de Agricultura Tropical shared data on the Colombian municipio boundaries. Malaria cases notified per capita were drawn from the reports of the Servicio Nacional de Erradicación de la Malaria (SEM) and refer to 1956. (Poveda, Graham and Epstein, 2000; Andrew Mellinger, private communication; Jonnes and Bell, 1997; SEM, 1957; and author’s calculations.)

To account for measurement error in the above variables, I also construct climate-based instruments. The set of instruments consists of the municipio’s temperature, altitude and the interaction of the two. The temperature and altitude data are from records prior to 1960. Another proxy for climate is the fraction of each municipio within particular Holdridge climate zones. Those relevant for the areas under study are the following: cool temperate, warm temperate, subtropical dry, subtropical wet, tropical dry, and tropical wet. These data come from a GIS file provided by the Center for International Development at Harvard University, and were computed by municipio in a GIS program (the “spatial join” in ArcView). (Banco de la República, 1964; Gallup, Sachs and, Mellinger, 1999b; and author’s calculations.)

- **Mexico.** Malaria mortality by state, expressed in per capita terms. (Pesquiera, 1957.)
Figure B – 1: Malaria Intensity by State in the United States

Notes: Displays a map of the ratio of malaria mortality to total mortality by state circa 1890. Source: Bureau of the Census (1894). Darker colors indicate more malaria.
Figure B – 2: Malaria Intensity by State in Brazil

Notes: Displays a map of an index of malaria ecology as constructed by Mellinger et al. (1998). Darker colors indicate climatic and geographic conditions more conducive to the transmission of malaria.
Figure B – 3: Malaria Intensity by Municipio in Colombia

Notes: Displays a map of an index of malaria ecology as constructed by Mellinger et al. (1998). Darker colors indicate climatic and geographic conditions more conducive to the transmission of malaria.
Figure B – 4: Malaria Intensity by State in Mexico

C Control Variables

Control variables for the United States:

- **Average wage, 1899.** I input the average monthly earnings (with board) for farm laborers by state in 1899. Various other wage measures are summarized by the same source, but are generally not available for a complete set of states. (Lebergott, 1964, Table A–24)

- **Region of birth.** These dummy variables correspond to the Census definition of regions: Northeast, South, Midwest, and West.

- **Doctors per capita, 1898.** Number of physicians per 1,000 inhabitants of each state. The primary source is listed as Polk’s *Register of Physicians*, 1898. (Abbott, 1900.)

- **State public-health spending, 1898.** Per capita appropriations, by state, for state boards of health in 1898. Primary sources include the annual reports of state boards of health, state appropriations laws, and correspondence with the secretaries of the boards of health. (Abbott, 1900.)

- **Child mortality, 1890.** The estimates of child mortality are constructed from published tabulations. Table 3 in Part III contains enumerated deaths of children under one year. I scale this number by the estimated birth rate (Part I, page 482) times the female population (Part I, Table 2). The rate from 1890 was used because child-mortality data are not available comprehensively for the years 1900–1932, during which time the death-registration system was established. The 1890 mortality data were collected by Census enumerators. (Census, 1894.)

- **Recruits for World War I found rejected for military service because of health “defects,” 1917-1919.** The fraction of recruits by state who were rejected by army physicians for physical defects. (Love and Davenport, 1920.)

- **Hookworm Infection.** Computed from examinations of army recruits. (Kofoid and Tucker, 1921)

- **Mortality from other diseases.** Separate variables are constructed for the following eight causes of death: scarlet fever, measles, whooping cough, diphtheria/croup, typhoid fever, diarrheal diseases, and pneumonia. Data are expressed as the fraction of total mortality in 1890. (Census, 1894.)

- **Fertility rate, 1890.** The estimated birth rate (from Part I, page 482). (Census, 1894.)

- **Log change in School Term Length, c. 1902–1932.** Average length of school term, in weeks. (Annual reports of the federal Commissioner of Education, 1905-1932.)

- **Log change in Average Monthly Salaries for Teachers, c. 1902–1932.** (Annual reports of the federal Commissioner of Education, 1905-1932.)

- **Log change in Pupil/Teacher Ratio, c. 1902–1932.** Average attendance divided by number of teachers. (Annual reports of the federal Commissioner of Education, 1905-1932.)

- **Log change in School Expenditure, c. 1902–1932.** (Annual reports of the federal Commissioner of Education, 1905-1932.)

- **Adult literacy rate.** These data were compiled at the state level and come from the 1910 Census. Adult literacy refers to males of voting age. (ICPSR #3.)

- **Population urban.** From Census tabulations measuring the population residing in metro areas in 1910. (ICPSR #3)

- **Fraction black.** From 1910 Census tabulations. (ICPSR #3)

- **Male unemployment rate.** From 1930 Census tabulations. (ICPSR #3.)

Control variables for the Brazilian states:

- **Region dummies.** North (Norte and Nordeste) and South (Centro-Oeste, Sudeste, and Sul).

• Infant mortality. Number of infant deaths in the municipio of the state capital, scaled by the estimated birth rate, which is computed from data for the whole state. (IBGE, 1951.)


• Fraction of population economically active. Measured for population ten years and older for 1950. (IBGE, 1950.)

• Shares of labor force by sector. Fraction of economically active population in each of the following sectors: agriculture, extractive industries, manufacturing, transportation, and services. Measured for population ten years and older for 1950. (IBGE, 1950.)

Control variables for Colombian municipios:

• Region dummies. The regions are as follows: Central, Bogota, Pacifico Norte, Eje Cafetero, Andina Norte, Andina Sur, Pacifico Sur, Caribe, Orinoquia, and Amazonia.

• “La Violencia”. A qualitative variable (ranging from 1 to 3) indicating the intensity of violence in the Colombian civil war known locally at “La Violencia”. Following Oquist, the coding of conflict intensity is split into subperiods: before 1955, when the violence was largely in population centers, and 1955 and after, when the conflict was more likely to take place in the countryside. (Oquist, 1976.)

• High Concentration “Minifundista”. Binary variable indicating the presence of small-land holders or minifundistas, as opposed to large land holders or urban areas. The reference period is the 1950s, although land-holding patterns were persistent historically. To construct municipio-level data, the map was digitized and georeferenced. Digital data on municipio boundaries, provided under special agreement from the Centro Internacional de Agricultura Tropical (CIAT), was overlaid on the map and municipios were coded dichotomously as indicated by the map. The municipio boundaries of the CIAT data refer to 1993, and therefore these mapped back onto 1950s entities. (Banco de la República, 1964 (map 57); Jonnes and Bell, 1997; DANE, 2001; Author’s calculations.)

• Coffee-growing Region. Binary variable indicating the presence of coffee cultivation. The reference period is 1960. Municipio-level data were created using the process described above for the “minifundista” variable. (Banco de la República, 1964, map 38.)

• Coal Mining Region. Dummy indicating the presence of actively exploited coal deposits, circa 1960. Municipio-level data were created using the process described above for the “minifundista” variable. (Banco de la República, 1964, map 22.)

• Expansion of Ranching. Areas identified for possible expansion of ranching in 1960. Municipio-level data were created using the process described above for the “minifundista” variable. (Banco de la República, 1964, map 55.)

• Infrastructure/Market Access. An index variable for the ease of transport to major markets or seaports from the area, based on infrastructure in circa 1960. Six (ordered) categories are used, following the map’s categorization. Municipio-level data were created using the process described above for the “minifundista” variable. (Banco de la República, 1964.)

• Level of development. An index variable for the general level of economic development of the area (“nivel de vida”), circa 1960. Six (ordered) categories are used, following the map’s categorization. Municipio-level data were created using the process described above for the “minifundista” variable. (Banco de la República, 1964, map 59.)

• Manufacturing employment per capita. Computed by municipio from the 1945 Colombian census of manufacturing. (Dirección Nacional de Estadística, 1947.)

Control variables for the Mexican states:
- **Region dummies.** “Norte” and “Sur”.

- **Population Density.** Population per square kilometer in 1950. (Dirección General de Estadística, 1952ab.)

- **Infant mortality.** Rate per 1,000 births. Data refer to 1950. (Coordinación General de los Servicios Nacionales de Estadística, Geografía e Informática, 1981.)

- **Log of Electricity Capacity.** Measured *circa* 1950. Original data in kilowatts. (Dirección General de Estadística, 1952b.)

- **Fraction of pop economically active.** Measured for population twelve years and older for 1950. (Dirección General de Estadística, 1952b.)

- **Shares of labor force by sector.** Fraction of economically active population in each of the following sectors: agriculture, extractive industries, manufacturing, transportation, and services. Measured for population twelve years and older for 1950. (Dirección General de Estadística, 1952b.)

- **Household income GINIs.** GINI coefficients were constructed from the 1950 census, which included tabulations by state of the distribution of families by household income. (Dirección General de Estadística, 1952b.)