

Tulane Economics Working Paper Series

The Long-Term Economic Impact of *In Utero* and Postnatal Exposure to Malaria

Alan Barreca
Department of Economics
206 Tilton Hall
Tulane University
New Orleans, LA 70118
abarreca@tulane.edu

Working Paper 0905
May 2009

Abstract

I use an instrumental-variables identification strategy and historical data from the United States to estimate the long-term economic impact of *in utero* and postnatal exposure to malaria. My research design matches adults in the 1960 Decennial Census to the malaria death rate in their respective state and year of birth. To address potential omitted variables bias and measurement-error bias, I use variation in "malaria-ideal" temperatures to instrument for malaria exposure. My estimates indicate that *in utero* and postnatal exposure to malaria led to considerably lower levels of educational attainment and higher rates of poverty later in life.

Keywords: early-life health, malaria, education, poverty
JEL: I12, I31, J0, J24

The Long-Term Economic Impact of *In Utero* and Postnatal Exposure to Malaria

Alan Barreca[*]

Tulane University

March 2009

Abstract

I use an instrumental-variables identification strategy and historical data from the United States to estimate the long-term economic impact of *in utero* and postnatal exposure to malaria. My research design matches adults in the 1960 Decennial Census to the malaria death rate in their respective state and year of birth. To address potential omitted-variables bias and measurement-error bias, I use variation in "malaria-ideal" temperatures to instrument for malaria exposure. My estimates indicate that *in utero* and postnatal exposure to malaria led to considerably lower levels of educational attainment and higher rates of poverty later in life.

[*] I am indebted to my dissertation chairs Hilary Hoynes and Douglas Miller for their help with this paper. Special thanks to Peter Lindert, Marianne Page, Ann Stevens, Elizabeth Cascio, Colin Cameron, Hoyt Bleakley, Alfredo Burlando, Trudy Marquardt, Janet Currie, as well as the participants at the UC Davis Brownbag series, the 2006 WEA Conference, and the 2007 SOLE Conference for providing valuable suggestions. This project was greatly aided by financial assistance from the Eugene Cota-Robles Fellowship and the Marjorie & Charles Elliott Fellowship awarded to me while I was attending the University of California, Davis.

I. Introduction

The argument that *in utero* and postnatal health conditions have a lasting impact on an individual's economic wellbeing rests on two premises: first, "early-life" health conditions predetermine one's health capital later in life; and second, health capital is an important determinant of economic capabilities throughout one's life.[1] The logic of this argument notwithstanding, the *extent* to which early-life health affects long-term economic wellbeing is unclear given the paucity of research on the subject. To help clarify the importance of early-life health conditions, this paper estimates the long-term economic impact of *in utero* and postnatal exposure to malaria. Exposure to malaria during this narrow timeframe has an acute impact on health. As such, understanding the consequences of such exposure will offer important insights into the significance of early-life health conditions over the life cycle.

The main challenge with estimating the effects of *any* early-life health shock (including exposure to malaria) is that omitted factors (e.g., socioeconomic conditions) are likely to bias observational studies towards overstating the importance of health. A number of recent studies have developed creative approaches to mitigating this potential omitted variables bias. For example, to test the long-term impact of unfavorable *in utero* health conditions, Almond (2006) compares long-term outcomes for cohorts born immediately subsequent to the 1918 influenza epidemic to the outcomes of cohorts born before or one year after the epidemic.[2] Meng and Qian (2006) test the long-term effects of early-life malnutrition as a result of being born during China's Great Famine (c. 1951-1961).[3] Also, Black *et al* (2005) and Oreopoulos *et al* (2006)

[1] The first premise originated with David Barker's "fetal-origins hypothesis" (Almond, 2006). See Barker (2001) for a review of the epidemiological evidence in support of this hypothesis.
[2] Almond (2006) finds large negative long-term consequences (e.g., lower income, lower educational attainment, higher disability rates) from exposure to the 1918 influenza epidemic.
[3] Meng and Qian (2006) find large negative effects (e.g., lower height, lower educational attainment) for cohorts who were exposed to China's Great Famine around the time of their birth.

estimate the long-term effects of *in utero* health conditions using differences between twins in order to control for observable and unobservable family-level factors.[4] However, the significance of early-life health conditions remains controversial because existing studies rely on a relatively limited set of identification strategies.

The key contribution of this paper is to develop a novel identification strategy to evaluate the long-term economic consequences of *in utero* and postnatal exposure to malaria. Specifically, I estimate the long-term effects of malaria using historical data from the United States and an instrumental-variables (IV) identification strategy. My research design matches outcomes of cohorts in the 1960 Census to the malaria death rate in their respective state and year of birth. To allay concerns of omitted-variables bias and measurement-error bias, I use within-state variation in "malaria-ideal" temperatures (i.e., between 22°C and 28°C) to instrument for malaria exposure at the time of birth. (As I discuss in greater detail in the Background section, these temperatures are favorable to the *Anopheles* mosquito, which is the vector for the malaria parasite.)

The United States experience with malaria presents an excellent opportunity to study the long-term effects of early-life health conditions for four reasons. First, malaria's physiological impact is acute during the *in utero* and postnatal periods relative to exposure at other ages.[5] Second, there is a large treatment group from which to estimate the effects of exposure since malaria was a serious public-health problem in the American South during much of the early 20th century. Conversely, cohorts born in states outside the South can be used as a control group

[4] Black *et al* (2005) and Oreopoulos *et al* (2006) both find that *in utero* health conditions positively affect long-term health, educational attainment, and labor-market outcomes.

[5] In general, contracting malaria, at *any age*, is harmful because the parasite destroys red blood cells when it reproduces in a human host's bloodstream, thereby depriving the body's tissues of oxygen and nutrients. For those exposed during the *in utero* and postnatal periods, the consequences of exposure to this disease are more acute because lowered immunity enables the malaria parasite to destroy relatively more red blood cells. An extensive epidemiological literature has documented the relationship between early-life exposure to the *falciparum* strain of malaria and short-term outcomes. (See Holding and Snow [2001] for a review of this literature.) Less is understood about the consequences from exposure to the *vivax* strain (WHO, 2007), which was the more prevalent strain in the United States during the early 20th century.

since it is unlikely they were ever exposed to malaria. Third, detailed vital statistics, climatic, and census data are publicly available for the United States. Fourth, the scope for selection bias is limited by the fact that malaria's fatality rate was relatively low in the United States.[6]

Moreover, my IV research design has several important features that enable me to address concerns that my instrument is correlated with important omitted variables. Using a repeated cross-section of state-year observations allows me control for time-invariant omitted variables, like geography, by including state fixed effects. I can also control for year-fixed effects and state-specific time trends to account for any spurious time-series correlations between climate changes and changes in birth conditions across state over time.

As a test of my exclusion restriction, I can include controls for temperatures that are ideal for agriculture as well as extreme temperatures that may be harmful to the human physiology. Also, I can test the effects of exposure to malaria-ideal temperatures in states where malaria was *not* a public-health problem as a test of my exclusion restriction. Similarly, by using the 1980 Census, I can estimate the effects of being exposed to malaria-ideal temperatures in states where malaria was once present *after* malaria was mostly eradicated.

In addition to contributing to the early-life health literature, the fact that malaria is a serious public-health problem in developing countries underscores the importance of my research. Cross-country estimates suggest that malaria-inflicted countries have significantly lower levels of income, all else equal (Gallup and Sachs, 2001). This paper is the first to estimate the long-term economic effects of *in utero* and postnatal exposure to malaria. As such, my work offers important insight into one of the potentially important mechanisms through which malaria might inhibit economic growth.

Furthermore, the development of my identification strategy in itself is a key contribution

[6] Approximately 200 to 400 cases were associated with each death (Humphreys, 2001).

to the malaria literature. Thus far, this literature has relied on malaria eradication campaigns or cross-sectional identification strategies to estimate the effects of malaria. For example, Bleakley (2003), Lucas (2005), and Bleakley (2006) find that malaria-eradication efforts led to long-term economic gains for children living in areas where malaria was once present.[7] Using a cross-section of predicted malaria incidence across the United States in the late-19th century, Hong (2007) demonstrates that there were significant long-term economic consequences for individuals who were born in malaria-inflicted counties.[8] My work complements the approaches used in these existing studies in that I rely on a different set of identifying assumptions.

The first-stage results indicate that daily mean temperatures between 22°C and 28°C are a strong predictor of malaria death rates within the United States during the early 20th century. For example, the estimated coefficient is statistically significant at the one-percent level with an F-statistic close to 10 in the core specification.

The IV estimates provide evidence that malaria exposure results in lower levels of income and higher rates of poverty. For example, the core IV estimate for log income indicates that exposure to an environment with 10 additional malaria deaths per 100,000 inhabitants causes a 13 percent reduction in income. Although my income estimates are not statistically significant at conventional levels, I find that malaria exposure has a statistically significant impact on the probability of being below 150 percent of the poverty level.

I also find that *in utero* and postnatal exposure to malaria significantly reduces

[7] Bleakley (2003) shows that malaria's decline in the United States led to moderate improvements in both long-term educational attainment and long-term earnings. Lucas (2005) examines the effects of malaria eradication campaigns in Paraguay, Sri Lanka, and Trinidad in the mid-20th century and finds that children born subsequent to malaria's decline made moderate gains in schooling. Bleakley (2006) illustrates that malaria eradication efforts in the United States, Brazil, Colombia, and Mexico led to long-term increases in income and literacy.

[8] Similar to my work, Hong uses weather to predict the incidence of malaria. However, my work has the added advantage of using high frequency (daily) temperature data, which allows me to more precisely identify the non-linear temperature-malaria relationship. While both Hong (2007) and Burlando (2007) share similarities with my research design, the work here was done in parallel to these other studies and the ideas developed independently.

educational attainment. For example, exposure to an environment with 10 additional malaria deaths per 100,000 reduces years of schooling by approximately 0.4 years, and reduces the probability of attaining at least 12 years of education by 5.1 percentage points (significant at the five- and ten-percent level, respectively). Furthermore, the magnitude of my schooling estimates suggest that *in utero* and postnatal exposure to malaria can account for as much as 25 percent of the difference in long-term educational attainment between cohorts born in malaria-afflicted states and cohorts born in non-afflicted states during the early 20th century.

Finally, there is some evidence that the long-term impacts are more severe for those exposed to malaria at the time of birth relative to those exposed to the disease in their second or third year of childhood. This result suggests that health policies targeting pregnant women and infants may have higher returns than those policies that target broader age groups.

II. Background on the relationship between temperature and malaria

Temperature affects the malaria life cycle in a variety of ways. Malaria is a parasite that can only be transmitted between humans via female *Anopheles* mosquitoes. These mosquitoes, which originate from larvae, first acquire the parasite when taking a blood meal from an infected human host. The malaria parasite then develops in the mosquito's stomach, a process called sporogeny. Thereafter the mosquito can infect a new human host when feeding again. The temperature-malaria relationship exists because larval development, mosquito survival, and sporogeny are all functions of the temperature.

Moreover, the temperature-malaria relationship is complex since larval development, mosquito survival, and sporogeny are each non-linear functions of temperature. For example, the proportion of larvae that reach adulthood (and become potential vectors for the malaria parasite)

is greatest when temperatures are between 20°C and 28°C (Bayoh and Lindsay, 2003).[9] Adult mosquito survival rates are relatively high until the temperature reaches 34°C; after which, mosquito survival rates decrease at an increasing rate until the probability of mosquito survival is nil at 40°C (Craig *et al*, 1999). Also, the rate at which sporogeny is completed is an increasing function of temperatures between 15°C and 40°C; consequently, most mosquitoes will not survive long enough to transmit malaria if temperatures are below 22°C (Craig *et al*, 1999).[10]

These numerous non-linear effects complicate efforts to model the temperature-malaria relationship. To simplify things, previous epidemiological studies have ignored much of the larval development stage of the malaria life cycle and instead have focused on modeling the probability that an adult mosquito would survive long enough to transmit malaria at least once.[11] Although the simplicity of these models is appealing, they overstate the importance of temperatures above the 28-30°C range since they neglect to fully account for larval development (Martens *et al*, 1995). Furthermore, some of their structural-form assumptions are questionable since they are derived from laboratory studies.

This paper follows a different tact than most epidemiological studies when modeling malaria transmission. I estimate the malaria death rate to be a function of the fraction of the year that the daily mean temperature is between 22°C (71.6°F) and 28°C (82.4°F), or the temperature "window" where malaria transmission is thought to be *least constrained* by temperature. As such, my model provides for easy interpretation and a transparent set of functional form assumptions. In addition to being grounded in epidemiology, below I show that the fraction of

[9] Also, the rate at which mosquito larva develop into adults is an increasing function of temperature between 16°C and 34°C (Bayoh and Lindsay, 2003).
[10] For example, sporogeny lasts approximately 210 days at 15°C, 14 days at 22°C, and 8 days at 30°C for the *vivax* strain of malaria. Sporogeny is possible between 15°C and 40°C for the *vivax* strain, and 16°C ad 40°C for the *falciparum* strain (Martens *et al*, 1995).
[11] For example, see Martens et al (1995) or Craig et al (1999).

the year with temperatures between 22°C and 28°C is, in fact, a strong predictor of the malaria death rate in the United States relative to other temperatures (see Figure IV).

Furthermore, some precipitation and/or the presence of permanent or semi-permanent bodies of water (e.g., lakes, swamps, streams) are necessary for malaria transmission because mosquito larva require standing water to survive until adulthood (Craig *et al*, 1999). This epidemiological facet of the malaria life cycle argues for using a model that interacts temperature with rainfall and/or interacts temperature with some proxy for the presence of standing water (e.g., fraction of the state's population living in a county with a swamp). Although I find some empirical evidence to support the inclusion of such interactions, my results are qualitatively similar without this added complexity.[12] As such, I model separable effects for temperature and rainfall for the sake of simplicity.

III. Empirical Methodology

A. Ordinary Least Squares (OLS)

My base OLS specification is of the following form:

$$(1)\ Y_{jt} = \beta \cdot \text{MALARIA}_{jt} + \gamma_j + \delta_t + \alpha_j \cdot \text{YEAR} + \varepsilon_{jt},$$

where Y is some average outcome (e.g., years of schooling, log of income) for the cohort born in state j and year t; MALARIA is the malaria death rate in state j and year t; γ is a set of state of birth fixed effects that account for any time-invariant differences between states (e.g., geography) that may also be correlated with MALARIA and Y; δ is a set of year of birth fixed effects, which account for any year-to-year changes in birth conditions that occur for the whole United States that are possibly correlated with MALARIA (e.g., business cycles); and $\alpha_j \cdot \text{YEAR}$

[12] The results from these specification checks are available upon request. However, I do include a temperature-precipitation interaction term in one specification below.

is a set of state-specific linear time trends that address the concern that the observed (downward) trends in MALARIA within some states is spuriously correlated with a gradual convergence in outcomes across states over time. To account for the possibility that the stochastic error terms (ε) are correlated within states, I cluster the standard errors on the cohort's state of birth.

Despite the inclusion of state fixed effects, year fixed effects, and state-specific linear time trends, estimates of β may be biased if MALARIA is: (a) correlated with important unobservable factors, and/or (b) a noisy proxy for the true incidence of malaria.[13] The omitted variables are likely to bias estimates of β towards overstating the adverse effects of malaria exposure since the disease is positively associated with poverty and poor public health conditions, while measurement error may bias the estimator towards zero (in the classical case); the net effect of these two biases is uncertain, *prima facie*.[14]

B. Instrumental Variables (IV)

In order to address the potential biases in OLS, I instrument the malaria death rate with the fraction of days with daily mean temperatures between 22°C and 28°C (hereafter "malaria-ideal temperatures") for the set of states where malaria was a serious problem. That is, I incorporate the fact that variation in malaria-ideal temperatures will have little or no effect on the malaria death rate in states where few human hosts carry the malaria parasite. In practice, I allow the effects of malaria-ideal temperatures to vary depending on whether a state has one of the 15 highest malaria death rates in 1920 according to Maxcy (1923).[15] These 15 states (hereafter "high

[13] Troesken (2004) notes that malaria deaths were sometimes misclassified as typhoid fever deaths, and vice versa, because some of the symptoms are similar (e.g., high fevers).
[14] Note that if the fatality rate from contracting malaria varies across states or over time, there may be non-classical measurement error.
[15] These 15 states include, in order of the state with highest malaria death rate to the 15th highest malaria death rate: (1) Mississippi, (2) Florida, (3) Arkansas, (4) Louisiana, (5) South Carolina, (6) Georgia, (7) Texas, (8) Alabama, (9) North Carolina, (10) Tennessee, (11) Oklahoma, and (12) Missouri, (13) Kentucky, (14) Virginia, (15) Illinois.

malaria states") all had at least one malaria death per 100,000 inhabitants and accounted for an estimated 95 percent of all malaria deaths that occurred in the United States in 1920.

Although malaria primarily afflicted the 15 high malaria states during the early 20th century, I include the remaining states and the District of Columbia (hereafter "low malaria states") in my sample in order to more precisely estimate the year-of-birth fixed effects while absorbing relatively less of the identifying variation.[16]

In sum, I propose the first-stage specification to be of the form:

(2) $\mathrm{MALARIA}_{jt} = \theta_1 \cdot \mathrm{TEMP2228}_{jt} \cdot \mathrm{HIGH} + \theta_2 \cdot \mathrm{TEMP2228}_{jt} + \pi \cdot \mathrm{X}_{jt} + \psi_j + \lambda_t + \tau_j \cdot \mathrm{YEAR} + \mathrm{vjt}$,

where MALARIA is the malaria death rate for the population of state j in year t, TEMP2228 is the fraction of year t that state j had a mean daily temperature between 22°C and 28°C, HIGH is an indicator variable for whether state j was one of the 15 high malaria states; ψ is a set of state fixed effects; λ is a set of unrestricted year fixed effects; $\tau_j \cdot \mathrm{YEAR}$ is a set of state-specific linear time trend; and *v* is a stochastic error term. Finally, X is a vector of weather-related control variables for state j and year t. Specifically, X consists of the average daily precipitation (PRCP) and the average of the daily precipitation squared (PRCPSQD) in year t and state j both as main effects and interacted with the high malaria state indicator.[17]

As shown in Table II below, my estimates of θ_2 indicate that the fraction of the year with malaria-ideal temperatures has no statistically significant or economically meaningful impact on the malaria death rate in low malaria states. As such, I impose the assumption that $\theta_2$equals zero

---

Also, my estimates are qualitatively similar when I redefine the set of high malaria states to include only the 12 most malaria states.

[16] My results are similar when the sample is limited to just the 15 high malaria states.

[17] I include the average of the daily precipitation squared since increases in precipitation may have diminishing returns at the day-level. For example, excessive rainfall may destroy mosquito breeding grounds (Craig *et al*, 1999). In an alternative specification (not reported) I control for the square of the average daily precipitation; the results are qualitatively similar.

in order to improve both the efficiency and the transparency of my identification strategy.[18] However, I do include both TEMP2228 and TEMP2228·HIGH as regressors in several reduced-from models to verify that malaria-ideal temperatures do not have an impact on long-term outcomes in states where malaria was not a serious problem.

I use equation (2), with θ_2 set equal to zero in my core specification, to predict the malaria death rates in equation (1). More formally, I estimate the following second-stage model:

(3) $Y_{jt} = \beta \cdot \widehat{MALARIA}_{jt} + \gamma_j + \delta_t + \alpha_j \cdot YEAR + \partial \cdot X_{jt} + \varepsilon_{jt}$,

where $\widehat{MALARIA}$ is instrumented with TEMP2228·HIGH. As in equation (1), this model includes a set of state fixed effects, unrestricted year fixed effects, and state-specific time trends.[19] However, I also include a vector of weather-related control variables (X). Specifically, I control for the average daily precipitation (PRCP) and the average of the daily precipitation squared (PRCPSQD) in high and low malaria states, respectively, since precipitation likely affects birth conditions irrespective of malaria (e.g., via crop yields). As discussed below, I also control for extreme temperatures as well as agriculture-ideal temperatures in some specifications.

C. Non-parametric temperature-malaria model

I also estimate a non-parametric model to verify that temperatures between 22°C and 28°C are a good predictor of the malaria death rate relative to other temperature ranges. Specifically, I estimate the following equation via OLS for the sample of high malaria states:

(4) $MALARIA_{jt} = \sum_r b_r \cdot TEMP_{jtr} + c \cdot X_{jt} + d_j + h_j \cdot YEAR + ejt$.

This model shares some similarities to equation (2) in that it includes a set of state fixed effects (d),

[18] The results are nearly identical when I do not impose the assumption that θ_2 equals zero. These results are available upon request.

[19] The state-specific linear time trends are important so as to control for the possibility that state-specific climatic trends may be spuriously correlated with a gradual convergence in outcomes across states over time. In order to allow for more flexibility, I also include a set of state-specific quadratic time trends in one specification as a robustness check.

state-specific time trends ($h_j \cdot YEAR$), and controls for precipitation and precipitation squared (X). However, I do not include year-fixed effects due to lack of statistical precision.[20] Also, TEMP now denotes the fraction of the year t in state j that is in one of five temperature ranges/bins (r); specifically, I estimate the effects of temperatures below 0°C, between 0°C and 8°C, between 8°C and 15°C, between 22°C and 28°C, and above 28°C, respectively, with temperatures between 15°C and 22°C as the omitted category.[21]

D. Tests of the exclusion restriction

There are two noteworthy mechanisms through which malaria-ideal temperatures may affect outcomes irrespective of malaria, which would bias my estimate of the causal impact of malaria exposure. First, malaria-ideal temperatures may be correlated with extreme temperatures that are generally unfavorable to the human physiology. For example, epidemiological evidence suggests that both "cold" weather (e.g., below 8°C) and "hot" weather (e.g., above 28°C) are a potential danger to human health.[22] To address these concerns, I control for the frequency of days with temperatures below 8°C, and the frequency of days above 28°C in some specifications.

Second, malaria-ideal temperatures are likely correlated with crop yields, which may indirectly affect human health conditions (e.g., via improved nutritional intake or additional farm income). For example, Deschênes and Greenstone (2007a) (hereafter DG) posit that agriculture product increases linearly as the temperatures increase between 8°C and 32°C. Schlenker and Roberts (2008) (hereafter SR) find that crop yields increase as the temperature increases between

[20] Although the point estimates are similar when I include year fixed effects in this model, the standard errors are nearly double.

[21] I have also estimated a model with round 5°C temperature bins (e.g., 20-25°C). The results are qualitatively similar. Furthermore, I rely on temperatures between 22°C and 28°C as opposed to a 1°C or 2°C wider temperature range (e.g., 20°C to 30°C) because of model fit. That is, models with temperatures between 22°C and 28°C have higher R-squared than models with a somewhat narrower or a somewhat wider temperature window. (These results are available upon request.)

[22] See Deschenes & Morretti (2007), Deschenes & Greenstone (2007b), and Barreca(2008).

8°C and some threshold value (e.g., 32°C for cotton); after which, crop yields begin to decline as the temperature increases. As such, my estimates of the causal impact of malaria may be biased since temperatures between 22°C and 28°C lie at the upper limit of "agricultural-ideal" temperatures.

To address the concern that malaria-ideal temperatures are correlated with crop yields, I include two sets of controls. First, I control for weather that is ideal for most types of agriculture, as defined by DG. Specifically, I construct this control by integrating the *daily mean* temperature between 8°C and 32°C during the "growing season"; all days with temperatures above 32°C are assigned a value of 24.[23] This control variable is known hereafter as "degree days (8-32°C)".

Second, since malaria-afflicted states are also cotton-producing states, I control for cotton-ideal temperatures as described by SR.[24] This control is constructed in a similar fashion to the degree days variable above, except I integrate the *fraction of the day* spent in each 1°C interval between 8°C and 32°C during cotton's "growing season" (following SR); importantly, I include a separate variable that measures the fraction of the growing season that was spent at or above 32°C.[25] These controls are known hereafter as "cotton-ideal temperatures".

As an additional test of my exclusion restriction, I examine the effects of malaria-ideal temperatures in low malaria states using a reduced-form model. Also, I estimate the relationship between malaria-ideal temperatures and adult outcomes for those cohorts born subsequent to malaria's demise (c. 1937-1956) using the 1980 Census.[26] However, urbanization, migration, and

[23] The "growing season" for most crops, according to DG, is between April 1 and September 30. See DG for a more detailed discussion.
[24] According to the 1940 Census of Agriculture (Minnesota Population Center, 2004), cotton constituted 25 percent of the total crop value in the South. In other regions of the United States, cotton output was close to nil.
[25] I linearly interpolate the fraction of the day spent in each 1°C bin based on the minimum and maximum temperature of the day. The "growing season" for cotton, according to SR, is between April 1 and October 31. See SR for a more detailed discussion.
[26] I use the 1980 Census since it is the last census in which an individuals' year of birth can be derived; (i.e., the quarter-of-birth variable is not reported in the 1990 or the 2000 Censuses).

the introduction of social-insurance and crop-insurance programs that occurred in the 1930s weakens the validity of the latter test.

E. Conceptual issues

There are three main conceptual issues and/or limitations with my identification strategy worth noting. First, I cannot discern the specific mechanism through which cohorts are affected by malaria. That is, I may be measuring the indirect effects of being born in a state where, for example, their parents (or neighbors) were infected.[27] Additionally, there may be peer effects, which would likely reinforce any direct effect of exposure, and/or other general equilibrium effects.[28] Given data limitations, I am unable to separate these different mechanisms.[29]

Second, my core identification strategy does not distinguish between *in utero* and postnatal exposure because the identifying variation is at the year-of-birth level. To disentangle the separate effects of *in utero* and postnatal exposure, I have estimated a reduced-form model using quarterly variation in malaria-ideal temperatures along with state by year by quarter of birth cells. The results of this model, which are available upon request, are too imprecise to offer any substantive conclusions.[30]

Third, there is some concern that my estimates may understate the effects of contracting malaria early in life because physically weaker fetuses and infants may die from the disease, and

[27] For example, malaria exposure may cause loss of parental income, which could indirectly affect birth conditions.
[28] For example, assuming contracting malaria diminishes human capital, the return to skill should increase within those cohorts exposed to malaria. As such, high-skill workers within treated cohorts would see an increase in earning potential, while low-skill workers would see a decrease in earning potential; the net-effect of which is uncertain, *prima facie*. Furthermore, these general equilibrium effects may also spill over into non-treated cohorts as well.
[29] This interpretation issue is not necessarily unique to my research design; (see Almond [2006], for example).
[30] To disentangle the separate effects of in utero and postnatal exposure, I have also estimated my core specification on each of the four different quarters of birth separately. In short, I find that the first and fourth quarter births are the most affected by malaria exposure; this suggests that both *in utero* exposure and postnatal exposure have significant long-term effects. However, my quarter-of-birth estimates are generally imprecise and I cannot rule out equal effects of exposure across each of the four quarters of birth in most cases. (These results are available upon request.)

consequently, may be excluded from my sample. To explore the scope of the selection effect, I estimate the effects of malaria on the log of the cohort size. There is little evidence to support the concern that my estimates are biased by a selection effect.

IV. Data

Malaria mortality data. The data on malaria deaths was constructed from various volumes of the *Mortality Statistics of the United States* from the National Center for Health Statistics (NCHS) archives. The NCHS reports malaria deaths at the state-year level from 1900 through 1941. Using historical population estimates from the Census, I construct malaria death rates by dividing the malaria mortality counts by the state-year population (in 100,000s). Although annual malaria data is available through 1941, I restrict my sample to the years 1900 through 1936 because malaria death rates declined significantly starting in 1937 for reasons unrelated to my identification strategy.[31] In addition, malaria deaths are not available from the NCHS for all state-years between 1900 and 1936 because states began reporting mortality statistics for the first time at different points over this period. The NCHS sample of malaria death rates total 1,147 state-year observations.

Climatic data. Daily minimum and maximum ground temperature and daily precipitation, by weather station, were obtained from the National Climatic Data Center (NCDC). Unlike the malaria mortality data, I have weather information for at least one weather station for all state-years from 1900 through 1936.[32] Over the period from 1900 through 1936, there are approximately 900 weather stations nationwide on average, and about 20 weather stations per state on average. The full sample of NCDC state-year observations total 1,813.

[31] According to Humphreys (2001) numerous factors lead to the demise of malaria; among other things, she cites urbanization, New Deal policies (e.g., Agriculture Adjustment Act), the Tennessee Valley Authority and their economic development efforts, and a general reduction in poverty as the major determinants of malaria's eradication.
[32] I use an unbalanced panel of weather stations to make use of all the available data.

I construct state-year weather variables from the daily weather-station data by taking an inverse-distance population weighted average of the weather-stations within 100 miles of a given state.[33] For example, I construct my measure of malaria-ideal temperatures (TEMP2228), the fraction of days within state j and calendar-year t that the daily mean temperature is between 22°C and 28°C, as follows:

(5) $\text{TEMP2228}_{jt} = \sum_{s=1}^{S_{jt}} \sum_{d=1}^{D_t} \text{DUM2228}_{sjtd} \cdot \omega_{sjt}$,

where DUM2228 is a dummy variable that takes a value of 1 if the mean temperature is between 22°C and 28°C at weather station s, on day d of year t.[34] D is the total number of days (i.e., 365 or 366) in year t. S_{jt} is the number of (reporting) weather stations in or within 100 miles of state j in year t. I weight all weather station observations by ω, an inverse-distance weight of the state-county population within 100 miles of the weather station. All weather variables (e.g., precipitation) are aggregated to the state-year level in a similar fashion to TEMP2228.

Census data. Data on adult outcomes was compiled using the 1960 Decennial Census. I create state and year of birth cell averages for those born between 1900 and 1936, or those who were approximately 24 to 60 years of age at the time of the 1960 Census.[35] Cohort outcomes include: years of schooling, attained at least 12 years of education, log income, is below 150 percent of the poverty line, worked 50 or more weeks last year, worked 35 or more hours per week last year, employed at time of census, fraction female, and log of cohort size.[36] The cohort cells are merged with the malaria and weather data by state and year of birth.

Data limitations. It is important to note that the geographic unit of observation is at the

[33] Weights are county population estimates from the 1920 Census.
[34] Mean temperature is determined by taking the average of the maximum and minimum temperature for that day.
[35] I exclude races other than white or black.
[36] In order to include all observations, I first set income to zero if it is reported as negative; I then take the log of income plus one and the log wages plus one. Following Almond (2006), I rely on the probability of being below 150 percent of the poverty line as a good indication of "poverty" status. Fraction female is an important indicator of *in utero* health since females are more likely to survive adverse *in utero* health conditions (Trivers and Willard, 1973).

state level due to data constraints. Malaria deaths are not available at the county level, nor are county of birth identifiers available in the 1960 Census. As such, it is not possible to use a finer geographic area (e.g., county).

Furthermore, to make maximum use of available data, both the OLS and IV specifications use an unbalanced panel of state-years for which I have malaria death rates. As a robustness check, I estimate a reduced-form model for the balanced panel of cohorts born between 1900 and 1936.

V. Descriptive Evidence and First Stage Results

As Figure I illustrates, malaria was a public health problem in the United States that was particular to the American South during the early 20th century. For example, the states with at least one malaria death per 100,000 inhabitants in 1920 were all located within (or adjacent to) the South. Moreover, Arkansas, Florida, Mississippi, and Louisiana all had between 30 and 40 malaria deaths per 100,000 inhabitants in 1920.[37] Since approximately 200 to 400 infections are associated with each malaria death, Figure I suggests that malaria may have infected as many as one in eight people in the most malaria-afflicted states. Both socioeconomic conditions (e.g., housing conditions, nutrition, education levels) and climatic differences between the South and the rest of the United States can likely account for the cross-region differences in malaria death rates (Humphreys, 2001).

In addition to the significant cross-state (or cross-region) differences in malaria death rates, there is also ample *within*-state variation in the malaria death rate over time. For example, Figure II illustrates the malaria death rate between 1918 and 1941 for the four "most" malaria-

[37] To put this figure in perspective, Kenya had approximately 60 malaria deaths per 100,000 inhabitants in 2002 (WHO, 2004).

afflicted states: Arkansas, Florida, Mississippi, Louisiana.[38] These four states all had malaria death rates that nearly doubled over a two- or three-year period between 1918 and 1941. Also, malaria rates were declining over this period in these four states.[39] As a reminder, I restrict my core sample to the years between 1900 and 1936 because malaria death rates began approaching zero starting in 1937.

Figure II also highlights the fact that much of the within-state variation in malaria death rates follows a similar pattern across states. For example, Arkansas, Mississippi, and Louisiana experienced similar a spikes in malaria death rates in 1933. Both socioeconomic conditions and weather conditions are potential candidates to explain this within-state variation because they are correlated within geographic regions over time.

Cursory evidence suggests that socioeconomic conditions cannot explain much of the year-to-year variation in the malaria death rate within malaria-afflicted areas of the United States. For example, malaria deaths increased during a time of relative economic prosperity (1928), declined after the onset of the Great Depression (1930), and surged again in the midst of the Great Depression (1933) in Florida and Mississippi (Figure II).[40]

Malaria-ideal temperatures, unlike socioeconomic conditions, can explain both the cross-state and the within-state variation in malaria death rates. Table I shows that high malaria states had approximately twice as many days per year between 22°C and 28°C as low malaria states on average. Conversely, low malaria states had over four times as many days below 8°C as high

[38] The NCHS does not report malaria deaths prior to 1918 for any of these four states.
[39] The downward trend in malaria death rates is correlated with general improvements in public health conditions (e.g., better understanding about how to prevent malaria transmission) and socioeconomic conditions (e.g. urbanization) as well as a concerted effort to eradicate malaria in the South (Humphreys, 2001). This trend also underscores the importance of including state-specific time trends in order to control for any spurious time-series correlations.
[40] Also, there is no statistically meaningful relationship between within-state changes in per capita income and within-state changes in the malaria death rate between 1929 and 1936. (I use state-year per capita income estimates from the Bureau of Economic Analysis, which does not have per capita income estimates for years prior to 1929.) These results are not reported.

malaria states. Also, the difference in precipitation levels between high and low malaria states is less stark. For example, high malaria states had 0.12 inches of precipitation and low states had 0.10 inches of precipitation on average per day between 1900 and 1936.

Moreover, Figure V demonstrates that malaria-ideal temperatures are positively correlated with the malaria death rate *within* the four most malaria-afflicted states. For example, changes in the fraction of the year between 22°C and 28°C mirror changes in the malaria death rate in Mississippi and Florida between 1928 and 1936. Consequently, using a within-state IV identification strategy to estimate the long-term effects of malaria exposure seems promising.

The OLS estimates in Table II indicate that the fraction of the year with daily mean temperatures between 22°C and 28°C, in the 15 high malaria states, is a strong predictor of the malaria death rate. For example, the column (1) specification (without any weather controls) indicates that a 10 percentage point increase in the fraction of the year with malaria-ideal temperatures, in high malaria states, causes an additional 2.9 malaria deaths per 100,000 inhabitants (33.3 minus 3.4 times 0.1). Also, the magnitude of the effect of malaria-ideal temperatures is large; for example, an increase of 2.9 malaria deaths per 100,000 inhabitants represents almost 30 percent of the average malaria death rate in high malaria states (i.e., 9.5 malaria deaths per 100,000 inhabitants). Furthermore, the estimated effect for high malaria states is statistically significant at the one-percent level. As expected, variation in the fraction of the year with malaria-ideal temperatures in low malaria states has no statistically distinguishable effect on the malaria death rate.

My estimates are qualitatively similar when I control for precipitation and/or precipitation squared.[41] Although somewhat smaller in magnitude, my estimates are robust to

[41] The coefficient estimates on precipitation and precipitation squared are not statistically significant at conventional levels, nor are they large in magnitude.

controlling for a temperature-precipitation interaction term or including state-specific quadratic time trends.[42] In addition, my results are similar when I restrict my sample to the set of high malaria states (results not reported).

My core specification, which is presented in column (6), relies on variation in malaria-ideal temperatures in high malaria states only (TEMP2228·HIGH) to predict the malaria death rate. That is, I drop TEMP2228 since the estimated impact of this variable is nearly zero across all specifications reported in Table II. (Although I include precipitation and precipitation squared in my core specification, my second stage results are qualitatively similar to these variables' exclusion.) Moreover, the fact that the F-statistic on TEMP2228·HIGH in my core-specification is approximately 9.6 mitigates "weak instrument" concerns.

Finally, Figure IV demonstrates that temperatures between 22°C and 28°C are a good predictor of the malaria death rate relative to several other temperature ranges (e.g., above 28°C). Figure IV, which presents the parameter estimates of the non-parametric temperature-malaria model, shows that a 10 percentage point increase in the fraction of the year between 22°C and 28°C causes the malaria death rate to increase by approximately 3.3 deaths per 100,000 inhabitants relative to temperatures between 15°C and 22°C. Moreover, the estimated coefficient on temperatures between 22°C and 28°C is statistically significant from zero at the one-percent level, while the estimated coefficients on the other temperature ranges are not significantly different from zero at conventional levels.

VI. Effects on Long-Term Economic Outcomes

[42] My temperature-precipitation variable is constructed by first interacting a 30-day moving average of an indicator variable that takes a value of one if the daily mean temperature is between 22°C and 28°C with a 30-day moving average temperature variable at the weather station-day level. Then, I aggregate the temperature-precipitation interactions to the state-year level using the weighting method described in the Data section.

A. Main results

Cohorts born in high malaria states have significantly worse long-term outcomes on average than those born in low malaria states, according to the 1960 Census (Table III). For example, cohorts born in high malaria states had 9.71 years of schooling while cohorts born in low malaria states had approximately 11.15 years of schooling, on average. High school completion rates were around 40 percent and 56 percent for high malaria and low malaria state cohorts, respectively. Income and earnings are approximately 25 percent lower for high malaria state cohorts. Furthermore, high malaria state cohorts also have a much larger black population (i.e., 20 percent versus 2 percent). The magnitude of the disparity in outcomes and characteristics illustrates the importance of identifying factors (such as variation in malaria-ideal temperature) that exogenously shift the risk of malaria exposure.

The OLS estimates indicate that there is little or no relationship between educational attainment and the malaria death rate, but there is a statistically-significant negative relationship between the malaria death rate and other important indicators of economic wellbeing (Table IV). For example, 10 additional malaria deaths per 100,000, which is just above the average malaria death rate in high malaria states, is associated with approximately 11 percent less income and a 0.8 percentage point increase in the probability of being below 150 percent of the poverty line. However, as mentioned above, the OLS estimates may not provide reliable estimates of the causal impact of malaria exposure on account of omitted variables and measurement error.

My core IV estimates provide extensive evidence that malaria exposure has a significant impact on educational attainment (Table IV). For example, the estimated coefficient on the malaria death rate is negative and statistically significant at conventional levels for years of schooling, probability of attaining 8 years of schooling, and probability of attaining 12 years of

schooling.[43] The estimated effects on the educational attainment variables are large in magnitude; for example, exposure to 10 additional malaria deaths per 100,000 inhabitants causes a cohort to have approximately 0.4 fewer years of schooling and a 5.1 percentage point less chance of having at least 12 years of education. Also, I find that malaria exposure has no adverse effect on the probability of attaining at least 16 years of schooling, which is consistent with the fact that malaria, like many diseases, impacted more economically disadvantaged populations.

Although my point-estimates suggest that malaria exposure adversely affects other measures of long-term economic wellbeing, large standard errors preclude making strong conclusions in most cases. For example, exposure to 10 additional malaria deaths per 100,000 inhabitants reduces income by approximately 13 percent; however, I cannot rule out a large *positive* effect on income.[44]

I do find statistically significant effects on the probability of being below 150 percent of the poverty line and on the probability of having worked 50 or more weeks. For example, exposure to 10 additional malaria deaths per 100,000 inhabitants causes a 3.8 percentage point increase in the probability of being below 150 percent of the poverty line (significant at the 10-percent level).[45]

Importantly, my IV estimates indicate that selection bias is not a significant concern. That is, the estimated impact of the malaria death rate on the log of the cohort size is small, positive,

[43] Note that the expected effect of malaria exposure on years of schooling is ambiguous *prima facie*. As noted by an anonymous referee, improved health raises both the returns to schooling (e.g., increases in work-life expectancy) and the opportunity cost of schooling (e.g., healthier individuals make better laborers).
[44] For example, the estimated effect of malaria exposure on income can be bounded between a 40 percent decrease and a 15 percent increase. Bleakley (2003) found that individuals exposed to malaria during childhood had approximately 15 percent less earnings. Given the imprecision of my wage and income estimates, I cannot discern to what extent *in utero* and postnatal exposure can account for Bleakley's findings.
[45] There is also a statistically significant effect on being below 100 percent of the poverty line.

and not statistically significant at conventional levels.[46]

B. Discussion

In sum, educational attainment is the most noticeable mechanism through which *in utero* and postnatal exposure to malaria affects long-term outcomes. My estimates indicate that malaria exposure can account for much of the differences in schooling between high malaria and low malaria state cohorts. For example, the difference in the malaria death rate was approximately 8.9 and the difference in years-of-schooling is approximately 1.4 years. Using my years-of-schooling point estimate from Table IV, I find that exposure to malaria can account for approximately 25 percent of the difference in years of schooling between cohorts born in high and low malaria states (8.9 times 0.04 divided by 1.4).[47] Furthermore, based on malaria's fatality rate, I estimate that one additional case of malaria translates into approximately 10 fewer years of schooling.[48]

Although the estimated impacts on schooling are large, they are generally consistent with previous malaria studies that examine the long-term effects of exposure during the broader time horizon of childhood. For example, Bleakley (2003) finds that the average cost of one year of exposure during childhood reduced schooling levels by approximately 0.05 years in the United States; my estimate is about eight times larger, suggesting that exposure during the *in utero* and postnatal period may be much more important than exposure during the "average" childhood year. Also, Lucas (2006) finds that a ten percentage point reduction in malaria during childhood increased years of schooling by about 0.10 years in Paraguay, Sri Lanka, and Trinidad.

[46] The estimated effect of malaria exposure on cohort size can be bounded between an 8 percent reduction and a 12 percent increase.
[47] The fact that I estimate a local-average-treatment effect limits the validity of this thought experiment, however.
[48] 0.04 divided by 400 times 100,000 equals 10. Since the magnitude of this effect is larger than a "typical" amount of schooling, this thought-experiment suggests that there may be peer effects or general equilibrium effects from being born in a cohort that was exposed to a significant amount of malaria.

According to my estimates, a ten percentage point reduction in malaria deaths within high malaria states (i.e., one death per100,000 inhabitants) would lead to approximately 0.04 more years of schooling, or four-tenths of Lucas' estimate.

Finally, my results suggest that measurement error may be biasing my OLS estimates more than any omitted variables since the IV estimates for years of schooling and probability of attaining at least 12 years of schooling are much larger (in absolute terms) than my OLS estimates.[49] This observation underscores the importance of using accurate indicators of malaria exposure (e.g., morbidity) and/or an instrumental-variables identification strategy when estimating the effects of malaria.

VII. Robustness checks

A. Controlling for omitted factors

As discussed in the Methodology section, I include several different sets of controls in order to address concerns that my instrument might be correlated with important omitted factors (Table V). In these specification checks, and for the remainder of the paper, I focus my attention on the following four outcomes: log income, probability of being below 150 percent of the poverty line, years of schooling, and the probability of attaining 12 years of education.

In column (1), I control for potentially harmful "cold" and/or "hot" temperatures, i.e., temperatures below 8°C and temperatures above 28°C. In columns (2) and (3), I include controls for "degree days (8-32°C)" and "cotton-ideal temperatures", respectively, in order to account for any changes in crop yields. In column (4), I include all the aforementioned controls. Finally, column (5) controls for state-specific *quadratic* time trends. My estimates are generally robust to the inclusion of these controls, although the estimates are imprecise when state-specific quadratic

[49] However, for years-of-schooling and for the probability of attaining at least 12 years of education, the difference is not statistically significant at conventional levels using a Hausman test that adjusts for clustered standard errors.

time rends are included.

B. Reduced-form estimates, pre- and post-eradication

As an additional test of my exclusion restriction, I estimate reduced-form models for cohorts born between 1900 and 1936 and for cohorts born between 1937 and 1956 using the 1960 Census and the 1980 Census, respectively. There are three noteworthy results from this reduced-form analysis: First, the estimates shown in Panel A (Table VI) for TEMP2228·HIGH and TEMP2228 suggest that malaria-ideal temperatures adversely affect educational attainment and the probability of being below 150 percent of the poverty line for those cohorts born in high malaria states *prior to* malaria's decline. Furthermore, my reduced-form estimates, which use the full panel of state-years, are consistent with my core IV estimates, which rely on an unbalanced panel.

Second, the estimates shown in Panel A for TEMP2228 suggest that malaria-ideal temperatures had mostly no effect on long-term outcomes for cohorts born in low malaria states during the 1900-1936 period. However, malaria-ideal temperatures may have actually led to a small improvement in income for those cohorts born between 1900 and 1936 in low malaria states. The income result is not a first-order concern because the magnitude of the effect is small and it suggests I am potentially underestimating the effects of malaria exposure.[50]

Third, the estimates shown in Panel B for TEMP2228·HIGH and TEMP2228 suggests that malaria-ideal temperatures had little or no effect on most outcomes for cohorts born in high malaria states *subsequent to* malaria's demise. There is a statistically significant and economically meaningful relationship between the probability of cohorts attaining at least 12 years of education and malaria-ideal temperatures. However, once I restrict my sample to the

[50] There is evidence to suggest that malaria-ideal temperature have a small "culling" effect in low malaria states.

cohorts born between 1941 and 1956, the estimated probability of obtaining 12 or more years of education is diminished and no longer statistically significant at conventional levels. This result can be explained by the fact that cohorts born between 1936 and 1940 were still exposed to modest levels of malaria (Faust and Hemphill, 1948).

C. Different samples

Table VIII separately presents the IV estimates for white males, white females, black males, and black females. My estimates are qualitatively similar for white males, white females, and black females. Although my point estimates for black males are generally smaller than for the other three samples, large standard errors preclude making any strong conclusions regarding the different effects of malaria exposure across these demographic groups.

I also test the effects of restricting to the sample of cohorts born in high malaria states. The qualitative conclusions are unchanged.

D. Exposure during first, second, and third years of life

I estimate a reduced-form model to test the effects of exposure to malaria-ideal temperature in the first, second, and third years of life in Table VIII.[51] Although statistically imprecise, these estimates indicate that exposure during one's year of birth (t) is more important than exposure during the second (t+1) and third (t+2) years of life.[52] This result suggests that health policies that target pregnant women and infants may have higher a return than those

[51] This thought experiment is complicated by the fact that malaria-ideal temperatures affect the probability of malaria transmission in subsequent years. Malaria-ideal temperatures affect this year's mosquito population, which will in turn affect next year's mosquito population and the propensity for malaria to be transmitted. In addition, because the *vivax* strain can survive in humans over the winter, more people with malaria in year t may affect the number of malaria hosts in year t+1 (Humphreys, 2001). Furthermore, I also tried estimating an analogous IV model using malaria death rates and malaria-ideal temperatures in year t, t+1, and t+3. The results were imprecisely estimated.

[52] This result is also consistent with Meng and Qian (2006), who show that the effects of early-life malnutrition is most pronounced for those who are exposed to China's Great Famine around the time of birth. Also, the results are qualitatively similar when I include five years of malaria-ideal temperatures as regressors.

polices that target somewhat broader age groups. However, statistical imprecision precludes making strong conclusions in this regard.

E. Other checks (not reported)

I have verified that my results are not sensitive to slight modifications in my identification strategy. Specifically, I have tried redefining "malaria-ideal" temperatures to be between 21°C and 29°C or between 20°C and 30°C. I have also tried restricting my sample to cohorts born between 1900 and 1930, and cohorts born between 1910 and 1936, in two separate checks. My estimates are robust to these modifications.[53]

VIII. Conclusion

Using historical data from the United States (c. 1900-1936) and an instrumental-variables identification strategy, this paper estimates the long-term effects from exposure to malaria during the crucial *in utero* and postnatal periods. After instrumenting malaria exposure with malaria-ideal temperatures, I conclude that cohorts exposed to more malaria in their year of birth have significantly lower levels of educational attainment as adults. I also find that malaria exposure led to a higher probability of being below 150 percent of the poverty level. Although imprecise, my estimates also indicate that malaria exposure has an adverse impact on income levels.

Moreover, I estimate that *in utero* and postnatal exposure to malaria can explain as much as 25 percent of the difference in educational attainment between those born in high malaria states and those born in the low malaria states during the early 20th century. The magnitude of this effect suggests there are significant economic ramifications over the life course from early-

[53] However, the estimates for the 1900-1930 sample are imprecisely estimated. I have also examined whether the effects of malaria-ideal temperatures vary over time. Given that malaria incidence was declining during the early 20th century (see Figures II), the negative correlation between malaria-ideal temperatures and long-term outcomes may be smaller for younger cohorts in the 1960 Census. However, this test does not offer any meaningful insight because the parameters are imprecisely estimated. (These results are available upon request.)

life exposure to adverse health conditions. In addition, my results indicate that protecting pregnant women and infants from malaria may be an important policy tool for improving economic outcomes in many developing countries today.

References

Almond, Douglas, "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy* 114 (2006): 672-712.

Barker, David, "The Malnourished Baby and Infant," *British Medical Journal* 60 (2001): 69-88.

Barreca, Alan, "Climate Change, Humidity, and Mortality in the United States," *University of California, Davis mimeo* (2008).

Bayoh, M.N., and S.W. Lindsay, "Effect of Temperature on the Development of the Aquatic Stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae)," *Bulletin of Entomological Research* 93 (2003): 375-381.

Black, S.E., P.J. Devereux, and K.G. Salvanes, "From the Cradle to the Grave? The Effect of Birth Weight on Adult Outcomes," *Quarterly Journal of Economics* 122 (2007): 409-439.

Bleakley, Hoyt, "Disease and Development: Evidence from the American South," *Journal of European Economic Association* 1 (2003): 376-386.

Bleakley, Hoyt, "Malaria in the Americas: A Retrospective Analysis of Childhood Exposure," *University of Chicago Mimeo* (2006).

Bureau of the Census, *State population estimates*, http://www.census.gov/popest/archives/1980s/80s_st_totals.html, (accessed 2006)

Bureau of Economic Analysis, *State Annual Personal Income*, http://www.bea.gov/regional/spi/, (accessed 2007).

Burlando, Alfredo, in-person meeting with author, January 2007.

Costa, Dora L., and Joanna N. Lahey, "Predicting Older Age Mortality Trends," *Journal of the European Economic Association* 3 (2005): 487-493.

Craig, M.H., R.W. Snow, and D. le Sueur, "A Climate-based Distribution Model of Malaria Transmission in Sub-Saharan Africa," *Parasitology Today* 15 (1999): 105-111.

Deschênes, Olivier, and Michael Greenstone (DG), "The Economic Impacts of Climate Change: Evidence from Agriculture Profits and Random Fluctuations in Weather," *American Economic Review* 97 (2007a): 354-385.

Deschênes, Olivier, and Michael Greenstone, "Climate Change, Mortality and Adaptation: Evidence from Annual Fluctuations in Weather in the U.S.," *NBER Working Paper* (2007b).

Deschênes, Olivier, and Enrico Moretti, "Extreme Weather Events, Mortality and Migration," *NBER Working Paper* (2007).

Faust, Ernest Carroll, and Faye M. Hemphill, "Malaria Mortality and Morbidity in the United States for the Year 1946," *Journal of the National Malaria Society* 7 (1948): 285-292.

Gallup, J.L., and J.D. Sachs, "The Economic Burden of Malaria," *American Journal of Tropical Medicine and Hygiene* 64 (2001): 85-96.

Hanigan, Ivan, Gillian Hall, and Keith BG Dear, "A comparison of methods for calculating population exposure estimates of daily weather for health research," *International Journal of Health Geographics* 5 (2006): 1-16.

Holding, P.A. and R.W. Snow, "Impact of Plasmodium Falciparum Malaria on Performance and Learning: Review of the Evidence," *American Journal of Tropical Medicine and Hygiene* 64 (2001): 68-75.

Hong, Sok Chul, "A Longitudinal Analysis of the Burden of Malaria on Health and Economic

Productivity: The American Case," *University of Chicago Mimeo* (2007).

Humphreys, Margaret, *Malaria: Poverty, Race, and Public Health in the United States*, (Baltimore, MD: The John Hopkins University Press, 2001).

Lucas, Adrienne M., "Economic Effects of Malaria Eradication: Evidence from the Malaria Periphery," *Brown University Mimeo* (2005).

Martens, Willem J.M., Louis W. Niessen, Jan Rotmans, Theo H. Jetten, and Anthony J. McMichael, "Potential Impact of Global Climate Change on Malaria Risk," *Environmental Health Perspectives* 103 (1995): 458-464.

Maxcy, K.F., "The Distribution of Malaria in the United States as Indicated by Mortality," *Public Health Reports* 38 (1923): 1125-1138.

Meng, Xin, and Nancy Qian, "The Long Run Health and Economic Consequences of Famine on Survivors: Evidence from China's Great Famine," *IZA Discussion Paper Series* (2006).

Minnesota Population Center, *National Historical Geographic Information System: Pre-release Version 0.1*. (Minneapolis, MN: University of Minnesota, 2004).

National Center for Health Statistics, *Mortality Statistics of the United States* (numerous volumes), (Washington D.C.: Government Printing Office.)

National Climatic Data Center, *Climate Research Data*, http://www.ncdc.noaa.gov/oa/climate/research/ushcn/daily.html, (Accessed 2006).

O'Malley, Claudia M., "The Biology of Anopheles Quadrimaculatus," *Proceedings of the Seventy-Ninth Annual Meeting of the New Jersey Mosquito Control Association, Inc.* (1992): 136-144. http://www.rci/rutgers.edu/~insects/mal5.htm (accessed 2009).

Oreopoulos, Phil, Mark Stabile, Randy Walld, and Leslie Roos, "Short, Medium, and Long Term Consequences of Poor Infant Health: An Analysis using Siblings and Twins," *NBER Working*

Paper (2006).

Rising, Bill, "SPHDIST: Stata module to compute spherical distances," Statistical Software Components S372502, Boston College Department of Economics, revised 19 June 2000 (1999).

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander, *Integrated Public Use Microdata Series: Version 3.0*, (Minneapolis, MN: Minnesota Population Center, 2004).

Schlenker, Wolfram, and Michael Roberts, "Estimating the Impact of Climate Change on Crop Yields: the Importance of Nonlinear Temperature Effects," *NBER Working Paper* (2008).

Sechrist, Robert P., "Basic Geographic and Historic Data for Interfacing ICPSR Data Sets, 1620-1983, United States," (Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2000).

Snow, R.W., C.S. Molyneux, E.K. Njeru, J. Omumbo, C.G. Nevill, E. Muniu, and K. Marsh, "The effects of malaria control on nutritional status in infancy," *Acta Tropica* 67 (1997): 1-10.

Teklehaimanot, Hailay D., Marc Lipsitch, Awash Teklehaimanot, and Joel Schwartz, "Weather-based prediction of Plasmodium falciparum malaria in epidemic-prone regions of Ethiopia I. Patterns of lagged weather effects reflect biological mechanisms," *Malaria Journal* 3 (2004): 1-11.

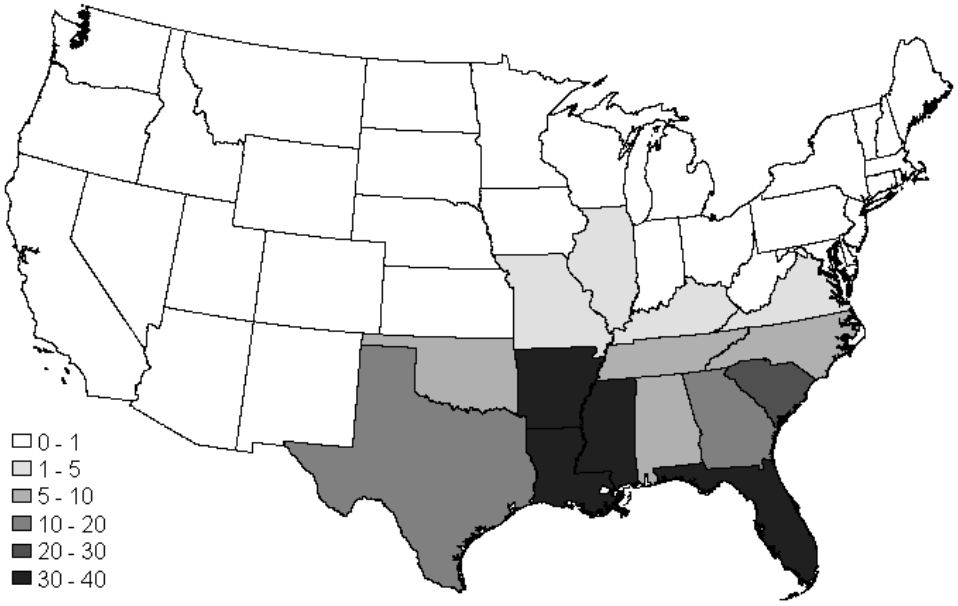
Troesken, Werner, *Water, Race, and Disease*, (Cambridge, MA: The MIT Press, 2004).

Trivers, Robert L., and Dan E. Willard, "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring," *Science* 179 (1973): 90-92.

World Health Organization, *Death and DALY estimates for 2002 by cause for WHO Member States*, http://www.who.int/entity/healthinfo/statistics/bodgbddeathdalyestimates.xls, (Accessed 2004).

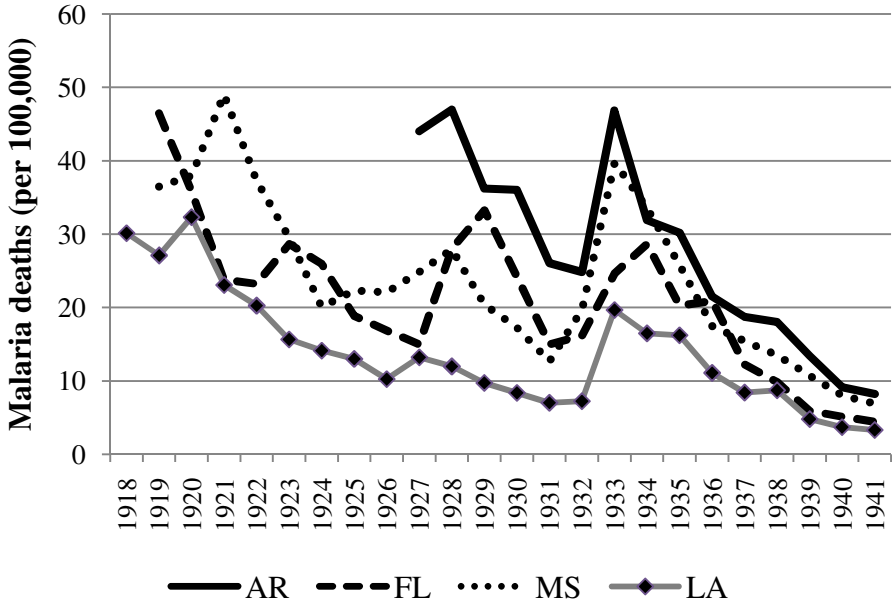
World Health Organization, *Malaria in Pregnancy*, http://www.rbm.who.int/cmc_upload/0/000/015/369/RBMInfosheet_4.htm (Accessed 2007).

Figure I: Estimated number of malaria deaths per 100,000 inhabitants in 1920, by state



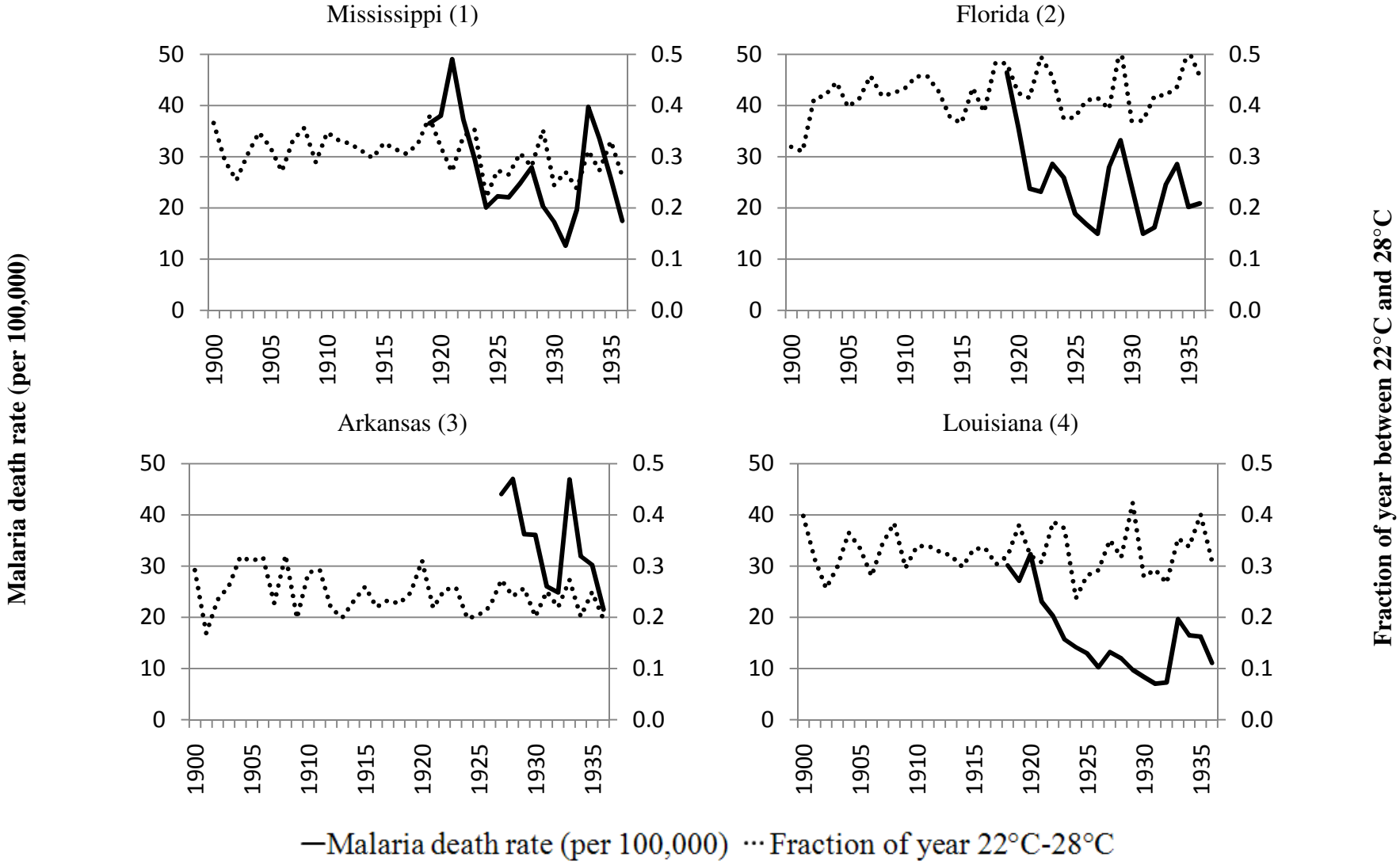
Data sources: National Center for Vital Statistics (various years), Maxcy (1923). Note that the NCHS does not report malaria death rates for many states in 1920. The estimated malaria death rates from Maxcy (1923) are used in those cases. All states that have at least one malaria death per 100,000 inhabitants are defined as "high" malaria states.

Figure II: Malaria deaths per 100,000 inhabitants, 1918-1941
The four "most" malaria-afflicted states



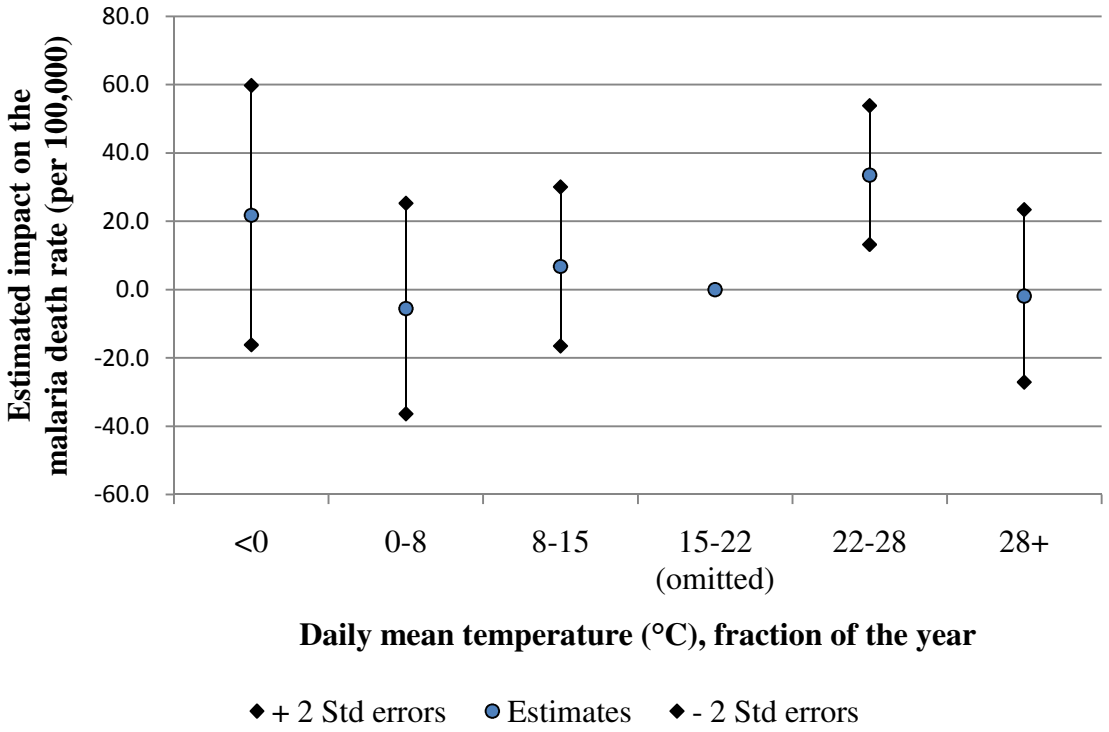
Notes: These states had the four highest malaria death rates out of all states in the United States in 1920 according to Maxcy (1923). I do not report years prior to 1918 since the NCHS does not report malaria deaths prior to then for any of these four states.

Figure III: Malaria death rate vs. temperatures between 22°C and 28°C, 1900-1936
The four "most" malaria-afflicted states, in order



Notes: These states had the four highest malaria death rates out of all states in the United States in 1920 according to Maxcy (1923).

Figure IV: Non-parametric parameter estimates, 1900-1936
Relative to temperatures between 15°C and 22°C, sample of "high" states only



Notes: The details of this regression can be found in the Methodology section in the "Non-parametric model" sub-section.

Table I: Summary of state-year malaria and weather data, 1900-1936

	Low malaria states		High malaria states	
	Mean	Std. dev.	Mean	Std. dev.
Malaria death rate (per 100,000)	0.593	1.231	9.472	9.785
Fraction of year below 0°C	0.178	0.077	0.056	0.059
Fraction of year 0°C - 8°C	0.230	0.049	0.156	0.063
Fraction of year 8°C - 15°C	0.205	0.058	0.199	0.031
Fraction of year 15°C - 22°C	0.249	0.046	0.251	0.035
Fraction of year 22°C - 28°C	0.126	0.047	0.255	0.064
Fraction of year 28°C - 34°C	0.013	0.018	0.084	0.061
Fraction of year above 34°C	0.000	0.001	0.000	0.002
Daily precipitation (inches)	0.101	0.029	0.124	0.029
Malaria state-year observations	882		265	
Weather state-year observations	1,258		555	

Notes: All calculations were weighted by the underlying number of census observations in a given state-year cohort cell.

Table II: First-stage estimates, 1900-1936
Dependent variable: annual malaria death rate (per 100,000)

	Specification					
	(1)	(2)	(3)	(4)	(5)	(6)
Independent variable	No controls	Precipitation	Precipitation squared	Temp-Prcp interaction	Quadratic trends	Preferred specification
TEMP2228·HIGH	33.25	31.71	31.68	23.35	21.00	27.63
	(11.11)***	(10.29)***	(10.38)***	(10.12)**	(7.37)***	(8.92)***
TEMP2228	-3.37	-5.09	-5.06	-6.77	0.32	Dropped
	(4.28)	(4.26)	(4.32)	(4.04)	(3.05)	
[F-statistic on TEMP2228·HIGH and TEMP2228]	[4.8]	[4.9]	[4.8]	[3.1]	[4.9]	[9.6]
Weather controls†						
PRCP	No	Yes	Yes	Yes	Yes	Yes
PRCPSQD	No	No	Yes	Yes	Yes	Yes
TEMP2228-PRCP interaction	No	No	No	Yes	No	No
Fixed effects						
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes	Yes
State-specific quadratic trends	No	No	No	No	Yes	No
Observations	1,147	1,147	1,147	1,147	1,147	1,147

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors (in parenthesis) are clustered on state. All regressions are weighted by the underlying number of census observations in a given state-year cohort cell. †All weather controls are allowed to vary by "high" and "low" states, respectively.

Table III: Summary of state-year cohort means, 1900-1936 cohorts (1960 Census)

	Low malaria states		High malaria states	
	Mean	Std. dev.	Mean	Std. dev.
Female	0.51	0.03	0.51	0.02
Black	0.02	0.03	0.20	0.14
School years	11.15	0.84	9.71	1.17
8+ years of school	0.92	0.07	0.75	0.14
12+ years of school	0.56	0.14	0.40	0.13
16+ years of school	0.10	0.03	0.07	0.03
Total income (2006$)	23,940	3,844	19,044	3,922
Total income < $6,000	0.35	0.04	0.39	0.04
Wage income (2006$)	19,460	2,769	15,449	3,143
Poverty status < 150%	0.22	0.07	0.39	0.11
Worked 35+ hours/week last year	0.55	0.05	0.52	0.05
Worked 50+ weeks last year	0.47	0.05	0.44	0.05
Employed at time of census	0.65	0.04	0.64	0.04
Underlying census observations	1,658,288		1,168,960	
State-year cohort cells	1,258		555	

Notes: All calculations were weighted by the underlying number of census observations in a given state-year cohort cell. Income and wages are in 2006 dollars. Also, the summary statistics were derived from the full sample of state-year cohorts (i.e., not restricted to malaria deaths reporting states).

Table IV: Effects of malaria exposure on a variety of outcomes, 1900-1936 cohorts
Independent variable: malaria death rate (per 100,000) at year of birth

	Specification			
	OLS		IV	
Dependent variable	Coeff.	Std. Error	Coeff.	Std. Error
School years	-0.004	(0.004)	-0.040	(0.018)**
Grade 8+ (x100)	-0.011	(0.045)	-0.423	(0.211)*
Grade 12+ (x100)	-0.075	(0.043)*	-0.514	(0.261)*
Grade 16+ (x100)	0.015	(0.020)	0.043	(0.088)
Log (income)	-0.011	(0.004)***	-0.013	(0.014)
Income < $6,000 (x100)	0.147	(0.058)**	0.117	(0.177)
Log (wages)	-0.011	(0.005)**	-0.014	(0.015)
Poverty status < 150% (x100)	0.079	(0.043)*	0.383	(0.223)*
Worked 35+ hours/week (x100)	-0.098	(0.053)*	-0.122	(0.215)
Worked 50+ weeks (x100)	-0.047	(0.047)	-0.347	(0.151)**
Employed (x100)	-0.116	(0.042)***	-0.081	(0.165)
Female (x100)	0.032	(0.028)	-0.177	(0.174)
Log (cohort size)	0.002	(0.002)	0.002	(0.005)
PRCP, PRCPSQD†	No		Yes	
Year fixed effects	Yes		Yes	
State fixed effects	Yes		Yes	
State-specific linear trends	Yes		Yes	
Observations	1,147		1,147	

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors (in parenthesis) are clustered on state of birth. All regressions are weighted by the number of underlying individuals in the state-year cohort. The IV estimates use TEMP2228·HIGH to instrument for the malaria death rate. †The IV estimates also control for precipitation and precipitation squared in "high" and "low" malaria states, respectively.

Table V: Instrumental-variables estimates with added controls, 1900-1936 cohorts (N=1,147)
Independent variable: malaria death rate (per 100,000) during the cohorts year of birth

	Specification				
	(1)	(2)	(3)	(4)	(5)
Dependent variable	"Hot", "cold"	Degree days	Cotton	All controls	Quadratic trends
School years	-0.068	-0.039	-0.047	-0.061	-0.045
	(0.027)**	(0.017)**	(0.016)***	(0.024)**	(0.026)*
Grade 12+ (x100)	-0.729	-0.518	-0.580	-0.647	-0.602
	(0.366)*	(0.253)**	(0.246)**	(0.334)*	(0.373)
Log (income)	-0.020	-0.013	-0.011	-0.019	0.007
	(0.020)	(0.013)	(0.014)	(0.019)	(0.018)
Poverty status < 150% (x100)	0.653	0.379	0.485	0.577	0.348
	(0.290)**	(0.219)*	(0.205)**	(0.260)**	(0.289)
Weather Controls†					
PRCP, PRCPSQD	Yes	Yes	Yes	Yes	Yes
TEMP < 8°C and TEMP > 28°C	Yes	No	No	Yes	No
Degree days (8-32°C)	No	Yes	No	Yes	No
Cotton-ideal temperatures	No	No	Yes	Yes	No
Year fixed effects	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes
State-specific quadratic trends	No	No	No	No	Yes
First-stage F-statistic	8.9	10.5	11.0	9.2	10.6

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors (in parenthesis) are clustered on state of birth. All regressions are weighted by the number of underlying individuals in the state-year cohort. These IV estimates use TEMP2228·HIGH to instrument for the malaria death rate. †The weather control variables are allowed to vary for "high" and "low" malaria states, respectively.

Table VI: Reduced-form estimates

	Dependent variable			
Independent variable	Years of schooling	12+ years of schooling (x100)	Log (income)	Poverty status < 150% (x100)
Panel A: Pre-eradication cohorts, 1900-1936 (1960 Census, N=1,813)				
TEMP2228·HIGH	-1.045	-14.784	-0.935	11.588
	(0.332)***	(4.641)***	(0.424)**	(4.244)***
TEMP2228	0.216	4.874	0.735	-4.77
	(0.298)	(4.490)	(0.422)*	(4.622)
Panel B: Post-eradication cohorts, 1937-1956 (1980 Census, N=980)				
TEMP2228·HIGH	-0.205	-10.715	0.338	1.553
	(0.437)	(5.259)**	(0.411)	(5.600)
TEMP2228	0.028	3.133	-0.049	2.223
	(0.338)	(3.801)	(0.347)	(3.839)
PRCP, PRCPSQD†	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors (in parenthesis) are clustered on state of birth. All regressions are weighted by the number of underlying individuals in the state-year cohort. †The reduced-form estimates control for precipitation and precipitation squared in "high" and "low" malaria states, respectively.

Table VII: Effects of malaria exposure on a variety of samples, 1900-1936 cohorts
Independent variable: malaria death rate (per 100,000) during the cohorts year of birth

	Sample				
Dependent variable	White males	White females	Black males	Black females	High malaria states
School years	-0.052	-0.036	-0.019	-0.041	-0.054
	(0.031)*	(0.026)	(0.042)	(0.035)	(0.021)**
Grade 12+ (x100)	-0.818	-0.455	0.000	-0.862	-0.671
	(0.489)	(0.419)	(0.429)	(0.462)*	(0.285)**
Log (income)	-0.011	-0.025	0.005	-0.014	0.017
	(0.017)	(0.032)	(0.023)	(0.037)	(0.015)
Poverty status < 150% (x100)	0.826	0.171	-0.242	0.497	0.335
	(0.483)*	(0.263)	(0.432)	(0.363)	(0.301)
PRCP, PRCPSQD†	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes
Observations	1,147	1,147	838	844	265

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors (in parenthesis) are clustered on state of birth. All regressions are weighted by the number of underlying individuals in the state-year cohort. †The IV estimates control for precipitation and precipitation squared in "high" and "low" malaria states, respectively.

Table VIII: Reduced-form estimates by year of exposure
1900-1936 cohorts (1960 Census, N=1,147)

	Dependent variable			
Independent variable	Years of schooling	12+ years of schooling (x100)	Log (income)	Poverty status < 150% (x100)
TEMP2228·HIGH (t)	-1.066	-14.599	-1.209	12.167
	(0.377)***	(5.667)**	(0.462)**	(4.472)***
TEMP2228·HIGH (t+1)	-0.291	-13.093	3.262	3.263
	(1.350)	(19.343)	(1.756)*	(14.390)
TEMP2228·HIGH (t+2)	0.603	16.737	-4.199	2.907
	(1.384)	(20.639)	(1.875)**	(14.637)
TEMP2228 (t)	0.156	3.918	0.617	-3.948
	(0.304)	(4.713)	(0.413)	(4.617)
TEMP2228 (t+1)	0.324	4.982	-1.511	-1.426
	(1.010)	(17.689)	(1.492)	(11.779)
TEMP2228 (t+2)	-0.840	-12.297	2.164	-3.734
	(1.081)	(18.763)	(1.595)	(11.376)
PRCP, PRCPSQD†	No	No	No	No
Year fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes
Observations	1,813	1,813	1,813	1,813

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors (in parenthesis) are clustered on state of birth. All regressions are weighted by the number of underlying individuals in the state-year cohort. †The IV estimates control for precipitation and precipitation squared in "high" and "low" malaria states, respectively.