Mosquitoes:
The Long-term Effects of Malaria Eradication in India

David Cutler Winnie Fung Michael Kremer Monica Singhal
Harvard University

This version: May 2007
Preliminary draft

Abstract

We examine the effects of malaria on educational attainment by exploiting geographic variation in malaria prevalence in India prior to a nationwide eradication program in the 1950s. We find that malaria eradication resulted in gains in literacy and primary school completion of approximately 10 percentage points. These estimates imply that the eradication of malaria can explain approximately half of the gains in these measures of educational attainment between the pre- and post-eradication periods in areas where malaria was prevalent. We do not see effects of the program in urban areas, where malaria was not considered to be a problem in the pre-eradication period. Women experienced smaller gains from malaria eradication than men, but we do not find disparities for members of scheduled castes and tribes.

1 Introduction

Malaria is one of the three infectious diseases that are responsible for the most deaths in human history. Today, malaria is endemic in over 100 countries and affects 40% of the world population. The World Health Organization (WHO) estimates that there are 300 million malaria cases and almost one million deaths from malaria in the world each year. Faced with this huge global burden, international organizations have redoubled their efforts to combat the disease. The United Nations has set combatting malaria as one of its Millennium Development Goals. The Roll Back Malaria Global Partnership, formed by the WHO, United Nations Children's Fund, United Nations Development Program, and the World Bank, aims to halve the malaria burden by 2010.

Many argue that improving health, while important in and of itself, can also lead to higher economic growth and development. Bloom and Sachs (1998) use cross-country regression analysis to argue that eradicating malaria in sub-Saharan Africa can increase that continent's growth rate by as much as 2% a year. Other macroeconomic studies, such as those by Gallup and Sachs (2001), Alleyne and Cohen (2002), and Bloom and Canning (2005), also conclude that improvements in health can lead to higher economic growth. However, it is not clear that these studies successfully isolate the impact of malaria from other factors correlated with malaria. Acemoglu and Johnson (2005) argue that the wave of international health innovations that began in the 1940s did not lead to a disproportionate increase in log per capita GDP or average levels of education in the areas subject to diseases affected by these conditions. Instead, in this paper, we use the malaria eradication program in India in the 1950s as a quasi-experiment to study the effects of malaria eradication on human capital accumulation. We exploit geographic variation in the prevalence of malaria prior to the eradication campaign and compare educational gains for cohorts born before and after the program in areas with varying pre-eradication prevalence. This study relates to several recent papers. Lucas (2005) looks at Paraguay, Sri Lanka, and Trinidad in the 1940s to 60s and finds that malaria eradication leads to increases in female education and literacy rate. Bleakley (2007) studies malaria eradication campaigns in the United States (circa 1920s) and in Brazil, Colombia, and Mexico (circa 1950s). He finds that childhood exposure to malaria lowers labor productivity which leads to lower adult income.

Two other studies use weather conditions to instrument for malaria exposure in the United

States and examine the effects on long run health and economic outcomes. Hong (2007) finds that malarial risk leads to adverse long run health outcomes, lower labor force participation and lower wealth. Barecca (2007) finds that in utero malaria exposure leads to lower educational attainment. Since these studies focus on variations in malaria that are driven by fluctuations in weather factors, the estimates may be capturing primarily the effects of epidemic malaria, rather than endemic malaria.

Our study examines the effect of malaria eradication in a single large country. The variation in pre-eradication prevalence is at the district level, allowing us to exploit quite localized variation in malaria exposure. It is the first to examine a sample containing both men and women, allowing us to examine gender disparities in long run impacts of the program. We also test for other sources of heterogeneity in treatment effects, to see whether traditionally disadvantaged groups also benefited from the program.

We find that the eradication program had a large, positive and significant effect on literacy and primary school completion. The effects are smaller but still significant for middle and secondary school completion. The results are robust to a variety of specification checks and cannot be explained by convergence across areas over time. We find smaller gains from malaria eradication for women than for men. Members of scheduled castes and scheduled tribes experienced similar gains to the rest of the population.

The paper proceeds as follows. Section 2 provides an overview of malaria prevalence in India and the malaria eradication program. Section 3 outlines our empirical strategy and Section 4 describes the data. Section 5 presents our results and Section 6 concludes.

2 Malaria in India

2.1 Overview

> "The problem of existence in very many parts of India is the problem of malaria. There is no aspect of life in this country which is not affected either directly or indirectly by this disease. It constitutes one of the most important causes of economic misfortune, engendering poverty, diminishing quantity and quality of food supply, lowering the physical and intellectual standards of the nation and hampering increased prosperity in

every way." - *John Sinton, Director of Malaria Survey of India, 1933*

Malaria is a protozoal infection transmitted to human beings by mosquitoes. The classic symptom of malaria is bouts of fever with spikes on alternative days. Headaches, malaise, fatigue, nausea, and anemia are also common. Severe forms of the disease result in organ failure, delirium, impaired consciousness, and generalized convulsions, followed by persistent coma and death.

Prior to the eradication program, malaria was considered to be the greatest health problem faced by India. Survey evidence estimates that immediately after partition in 1947, India suffered from 75 million cases of malaria (doubled during epidemic years) and 800,000 deaths directly attributable to malaria annually. The population of India in 1947 was 344 million, implying an annual prevalence rate of 22%. The estimate of malaria deaths implies that malaria was responsible for approximately 10% of total annual deaths in the pre-eradication era.[1]

2.2 The Pre-Eradication Era

References to malaria can be found in Vedic writings dating to 1600 B.C. (Desowitz 1991) and two classical books on Ayurveda describe malaria as the "king of diseases" (NMCP 1986). Efforts to control malaria date back to the early 1900s. Early experiments prior to 1910 focused on breeding control. These attempts were generally considered failures. From 1910 to 1944, measures such as drainage and filling up of breeding places were undertaken. Some limited success was also achieved using larvicidal chemicals such as oil and Paris green, and later pyrethrum. Malaria control was revolutionized in the mid-1940s with the advent of DDT (dichlorodiphenyl trichloroethylene). DDT was effective, non-toxic to humans, and "dirt-cheap to manufacture" (Desowitz 1991). Aggressive campaigns using DDT were launched almost simultaneously around the world, leading to the rapid eradication of malaria in Taiwan, much of the Carribean, the Balkans, parts of northern Africa, northern Australia, and large parts of South Pacific (Davis 1956).

DDT was first used in India by the military in 1944 and became available for civilian anti-malaria operations in 1945. During the late 1940s, a number of pilot programs and trials took

[1] Of the four human malaria parasites (*Plasmodium falciparum, P. vivax, P. malariae* and *P. ovale*), two are prevalent in India: *P. vivax* and *P. falciparum.* *P. malariae* also exists, but is confined to tribal areas of the country (NMEP 1986). Data on the relative prevalence of these vectors in the pre-eradication era is unfortunately unavailable; data from the immediate post-eradication period suggests that approximately 30% of cases were due to *P. falciparum* (NMEP 1996). This percentage has increased in recent years. *P. falciparum* is associated with the most severe forms of malaria and accounts for most malaria fatalities.

place throughout the country. International organizations such as the WHO, UNICEF and the Rockefeller Foundation also sponsored demonstration projects. These pilot projects were very successful, and in 1951, the national Planning Commission endorsed the development of a comprehensive, nationwide malaria control program. In April of 1953, the National Malaria Control Program (NMCP) was launched. The funding for the program was primarily from bilateral and international sources, and the implementation of the program was overseen by the WHO. The timing of the program is plausibly exogenous, since it was driven by the advent of DDT.

2.3 National Malaria Control Program

The program's main operational activity was to spray DDT in human dwellings and cattle sheds. The spraying was done usually between May and September (peak transmission times). Two rounds of spraying were conducted during these months.

A five year plan was formed under which the target was to establish 125 Malaria Control Units, each covering a population of 1 million by 1956 (NMEP 1986). The program was successful in achieving this goal (*Table 1*). By 1956, 134 units had been established, and 112 million people were estimated to be protected. By 1958, almost 200 units had been established and 165 million people were under protection.

The program was so successful that in 1958 it was reformulated into the National Malaria Eradication Program with the goal of completely eradicating malaria from the nation. This effort was adopted in coordination with a WHO campaign to eradicate malaria from the entire region, launched after the Eighth World Health Assembly in 1955 (WHO 1967). By 1960-61 the entire country was brought under the program.

Figures 1.1-1.4 illustrate the rapid geographic expansion of coverage as districts were phased into the program. Once a district was put under the program, it was under the program in all subsequent years. The statement of the Planning Commission indicated that priority targeting of areas should be based on endemicity and food producing capacity. The timing of the phase-in for particular districts may therefore not be exogenous.

Malaria was not as prevalent in urban areas prior to the eradication era. In fact, urban malaria was considered to be a negligible problem and malaria control efforts were left to local governments. Prevalence of malaria in urban areas increased only later (although not nearly to the levels of pre-eradication rural malaria), and the government launched an Urban Malaria Scheme in 1971 to address the growing problem of urban malaria.

2.4 Effectiveness of the Program

While the campaign was unsuccessful in eradicating malaria from India, it did achieve tremendous reductions in malaria prevalence relative to the pre-eradication era. The NMEP, which began tracking malaria prevalence from 1961 using blood smear data, estimates that in 1965 there were 100,000 cases of malaria, as compared to 75 million cases and 800,000 deaths in the pre-eradication era (*Table 2*).

The 1965 figures almost undoubtedly underestimate true malaria prevalence; however, there is no doubt that there was a dramatic reduction in malaria prevalence over this period. *Table 3* shows three measures of malaria prevalence during the control program: the child spleen rate, the child parasite rate and the infant parasite rate. The child spleen rate is the percentage of examined children ages 2-9 with an enlarged spleen. The child or infant parasite rate is the percentage of children ages 2-9 or infants examined in a blood survey who are found positive for a malaria parasite. All three measures show substantial declines over this period. Vital statistics data on causes of death indicate that the number of malaria deaths between 1951 and 1963 dropped by 93.4% and 98.5% in Uttar Pradesh and West Bengal, respectively (*Figure 2*). These states were chosen because they are large states with relatively complete data and their boundaries did not change much during the reorganization of Indian states in 1956.

Malaria prevalence remained low throughout the 1960s but experienced a slight resurgence in the 1970s, peaking in 1976. However, even at the peak of the resurgence, the prevalence rate was only 1.1% (*Table 2*). Reported prevalence decreased again, although not to the low levels seen in the immediate post-eradication period. This may be partially a result of increased accuracy in reported caseloads over time.

3 Empirical Strategy

Our study focuses on the effects of malaria exposure on educational attainment. Malaria has been hypothesized to have lifelong effects on cognitive development and educational attainment through three channels: chronic malaria-induced anemia, time lost or wasted in the classroom due to illness, and low birth weight. Infants and children are the worst sufferers of malaria and are subjected to high morbidity and mortality. The other at risk group is pregnant women, in whom malaria often results in pre-term labor or low birth weight full-term births as well as spontaneous abortions and still-births. Epidemiological studies have associated malaria with anemia, epileptic convulsions, and growth faltering during the first three years of life (Shiff et al. 1996). These in turn can cause learning disabilities and affect cognitive development of children (Boyle et al. 1994).

Malaria can also affect school attendance. In Kenya, primary school students have been estimated to have on average four episodes of malaria per year and to miss five school days per episode, amounting to 20 school days missed per child per year (Leighton and Foster 1993). A study in The Gambia shows that the use of insecticide-treated mosquito nets, which presumably reduces the threat of malaria, lowers school absenteeism due to fever (Aikins 1995).

Lastly, malaria in pregnancy can cause low birth weight because of fetal growth retardation or premature delivery (Duffy and Desowitz 2001). This can in turn reduce the physical, cognitive, and neurosensory development of the child resulting in lower human capital accumulation (McCormick et al. 1992).

Since malaria likely affects cognitive development and educational attainment mainly during the infancy and childhood years, we focus on the effects of malaria exposure at birth. Malaria prevalence at birth can be seen as an approximation of the individual's malaria exposure during the first few years of life, when the effect of malaria is likely the strongest. We also report results adjusting the treatment and control windows.

Although there is variation in when districts were phased into the program during the eradication era, the timing of phase-in may be related to malaria severity and other relevant factors. In addition, measures of malaria prevalence (child spleen rate, child parasite rate and infant parasite rate) show declines in both sprayed and unsprayed areas over this period, suggesting that even those in unsprayed areas may have benefited from the program (NMCP 1986). Finally, the phase-in of the program was quite rapid, raising further difficulties in exploiting variation in timing of coverage.

The basis of our empirical strategy is a differences-in-differences design, exploiting geographic variation in the prevalence of malaria prior to the eradication program. We compare outcomes at a point in time for individuals in birth cohorts born before and after the eradication era in areas with high and low pre-eradication malaria prevalence. Ideally, we would like to know the district of birth for each individual; however, our outcomes data reports only the district of current residence. An identifying assumption of our analysis is therefore that district of residence is a good proxy for district of birth. Only 11% of individuals in the 1991 Census are reported as living in districts other than their district of birth.

We run regressions of the following form, for individual i in birth cohort c in district d:

$$Outcome_{icd} = \beta_0 + \beta_1 (Post)_c * (High)_d + \gamma_d + \alpha_c + X\beta_2$$

where $Post$ indicates that the individual's birth cohort is after the eradication era and $High$ indicates whether individual i's district was a high prevalence district prior to eradication. γ and α are district and birth cohort fixed effects, and X is a vector of individual characteristics, including sex, scheduled caste, scheduled tribe, and household religion.

4 Data

4.1 Map of Pre-Eradication Endemicity

A central problem in assessing the impact of malaria is the identification of a suitable indicator for the prevalence of the disease. As Gallup and Sachs (2001) point out, there is a lack of high-quality data on malaria incidence in the most severely affected countries. In their study, they use historical maps of the geographical distribution of malarial risk to derive an index of malaria prevalence.

In this paper, we use a 1948 government map that classifies areas into categories of malaria prevalence. The map was obtained from the Ministry of Health and Family Welfare, Government of India. The pre-eradication malaria map classifies areas into six categories of endemicity: (1) areas above 5000 feet; non malarious, (2) known healthy plain areas; spleen rate under 10%, (3) variable endemicity associated with dry tracts; potential epidemic areas, (4) known areas liable to fulminent epidemic diluvial malaria, (5) moderate to high endemicity; fulminent epidemics unknown, and (6) hyperendemicity of jungly hill tracts and terai land. This map was based on spleen rate surveys and climate factors, although the exact mechanism by which category boundaries were constructed is not known.

Using geographic information system (GIS) software, we digitize the 1948 malaria endemicity map. *Figure 3* shows the digitized map. There are 466 districts in the 1991 census. However, the National Sample Survey (NSS) coding groups some districts together, resulting in 431 NSS "districts." We follow the NSS district coding. Map prevalance data is unavailable for three island districts (Andaman & Nicobar Islands, and Lakshadweep) so we drop these districts. In our analysis, we group districts in categories 1 and 2 together, and define them as non-malarious regions. Districts in categories 3 and 4 are grouped together and defined as potential epidemic areas, while districts in categories 5 and 6, where malaria is endemic, are classified as malarious areas. We classify each district into one of these three groups.

As can be seen in *Figure 3*, some districts have more than one possible classification. 295 districts (68.9%) are "unambiguous"; that is, they contain only one of the six categories or two categories that fall in the same group (e.g. categories 1 and 2). 53 districts (12.4%) are "majority" districts. These districts have one group that comprises a majority of the possible categorizations and are assigned that group. For example, districts with categories 3, 4 and 5 are categorized as potentially epidemic. The remaining 80 districts (18.7%) have no majority and are "discretionary." In our main specifications, we assign these districts to the highest possible prevalence category. For example, districts with categories 4 and 5 are classified as malarious. Under this classification, 24 districts (5.6%) are coded as non-malarious, 114 (26.6%) are potential epidemic, and 290 (67.8%) are malarious. Misclassification of districts should, in general, bias our results downward. We also perform extensive robustness checks to test whether the method of district classification affects the empirical findings.

4.2 Outcomes

We use data from the 43rd round of the Indian National Sample Survey (NSS), conducted in 1987. The NSS is an all-India representative household consumer expenditure survey set up by the Government of India in 1950. It includes a parallel employment and unemployment survey every five years. The NSS has four "thick" rounds that have the largest samples: namely, 1983, 1987, 1993, and 1999. We use the 43rd round (1987) because it is the earliest thick round that contains district identifiers. Choosing an early round mitigates possible mortality bias, and using the district identifiers allows us to examine outcomes at a very local level. The NSS reports district of current residence but not district of birth.

Schedule 10 (employment schedule) of the NSS gives information on education and income. In the NSS, individuals report the highest level of education they have attained. The variable "literate" is a dummy variable that equals one if the individual can read and write. "Primary" is a dummy that equals one if the individual has at least finished primary school.

Table 4 provides summary statistics for the entire survey sample in the 43rd round. Almost 300,000 households were sampled, giving slightly over 1 million individual observations. 67% of sample individuals live in rural areas, 40% are female, and the average age in the sample is 28.

For our main specifications, we focus on the rural sample. We restrict the sample to individuals aged 15 and over for literacy and primary school completion outcomes and to aged 20 and over for higher level educational outcomes. We exclude those born during the eradication era (1953 to 1961). The ten birth cohorts born from 1962 to 1972 are considered post-eradication cohorts and the forty birth cohorts born from 1912 to 1952 are considered pre-eradication cohorts. We demonstrate in the robustness analysis that the results are not sensitive to the choice of age cutoff or the choice of initial birth cohort for the pre-eradication group.

Table 5 provides summary statistics for the pre- and post-eradication cohorts. For the pre-eradication cohorts, the average literacy rate and completion rate for all levels of schooling is highest in the non-malarious districts, consistent with malaria prevalence depressing educational attainment. The differences in educational outcomes between the non-malarious districts and the potentially epidemic and malarious districts are generally much smaller for the post-eradication cohorts. In some cases, the point estimates for educational attainment variables are even higher for the districts that were malarious in the pre-eradication period relative to those that were non-malarious.

5 Results

5.1 Baseline Results

Table 6 shows the results of our baseline specification for a number of educational and income outcomes. The coefficients of interest are the coefficients on post*malarious and post*potential epidemic, which capture the effect of being born during the post-eradication period in a district that was malarious or potentially epidemic in the pre-eradication period. The OLS estimates for literacy rate, primary, middle, and secondary schooling are positive and statistically significant, reflecting the benefits of malaria eradication on educational attainment. The benefits decrease in magnitude for the higher levels of education completion. Effects on college education are insignificant. This is perhaps unsurprising, given that the nationwide average level of college completion is less than 2%, even in the post-eradication era.

Column (1) shows that the gain in literacy rate between those born in post-eradication versus pre-eradication cohorts is 9.3 percentage points higher for those born in a malarious region compared to those born in a non-malarious region. Those born in potential epidemic areas also benefit from a 8.5 percentage point relative increase in literacy rate. Column (2) shows that malaria eradication leads to a 10.1 percentage point increase in the probability of completing at least primary school for those born in malarious regions, and a 9.0 percentage point increase for those born in potential epidemic areas. The benefits decrease in magnitude for the higher levels of education completion (middle and secondary school) but are still significant at the 10% level. The gain in all educational attainment variables is higher for those in malarious areas as compared to potential epidemic areas, as we would expect since malarious areas experienced larger declines in malaria prevalence.

These increases are quite substantial, relative to a nationwide average literacy rate of 32% and a primary school completion rate of 20% for the pre-eradication cohorts (*Table 5.1*). If we compare column (1) in *Table 5.1* and *5.2*, we see that the raw difference in literacy rate between the pre- and post-eradication cohorts in malarious regions is 0.505 - 0.331 = 0.174. This means malaria eradication accounts for (0.093 / 0.174) = 53.4% of the increase in literacy rate in malarious regions. Similarly, malaria eradication accounts for (0.085 / (0.456 - 0.268)) = 45.2% of the increase in literacy rate in potential epidemic areas. Similar calculations demonstrate that eradication was responsible for 51.8% of the increase in primary school completion in malarious areas and 42.7% of the increase in potential epidemic areas.

There are several things to note when interpreting the above results. First, cohort attrition may lead to a downward bias in the estimated effect of malaria eradication. In the pre-eradication era, there were 75 million cases of malaria per year and 800,000 deaths. Most of the deaths were of children under the age of five. Those who survived to older ages were presumably the stronger cohort members. Weaker cohort members who would have died in the pre-eradication period were able to survive past eradication. Hence the educational attainment of individuals in the eradication period is averaged over both strong and weak cohort members, and the estimated gain in educational attainment for those who would have survived is therefore likely to be downward biased.

Second, the estimated effect of malaria eradication includes both direct and indirect effects. An example of a direct effect is the reduction in school absence for an individual who now does not suffer from malaria. An example of an indirect effect is the reduction in school absence for an individual who now does not need to take care of younger siblings who suffer from malaria.

5.2 Robustness Tests

5.2.1 Empirical Specification and Sample Selection

Our robustness analysis focuses on the literacy and primary school completion outcomes, since these effects are the strongest in the baseline specification. *Column 2* of *Table 7* presents the results with district and year of birth dummies but excludes individual control variables. Excluding the individual controls causes the point estimates to decrease slightly, but the estimated effects are still quite close to the baseline effects, and all are statistically significant at or above the 10% level. *Column 3* excludes Jammu and Kashmir as well as Himachal Pradesh, to make sure that the results are not due only to these Himalayan areas, which could have been subject to other factors. Our results are slightly stronger when these states are excluded.

We next adjust the choice of pre- and post-eradication cohorts. First, we limit the pre-eradication cohorts to the 10 birth cohorts born between 1942 and 1952. Second, we use a broader definition of the eradication era. We have in effect assumed that those born before 1953 have no benefit from the program and those born after 1961 are fully covered. We therefore redefine the eradication era to be the birth cohorts from 1950 to 1963; the pre-cohorts are the 1940-49 birth cohorts, and the post-cohorts are the 1964-1973 birth cohorts. Finally, we restrict the age cutoff to be those individuals aged 20 or over, which then defines the post-eradication cohorts to be those cohorts born between 1962 and 1967. These results are presented in *Columns 4 to 6*. Varying the pre- and post- windows does not affect our results. All of the effects continue to be statistically significant, and the point estimates are very similar to the baseline results. Restricting the pre-eradication cohorts to the 1942-1952 cohorts and restricting the age cutoff to 20 or above decrease our point estimates for both literacy and primary school slightly; expanding the definition of the eradication era causes the point estimates to increase slightly. All of the estimated effects are within a few percentage points of the effects in the baseline specification.

We also examine the effects on the eradication era cohorts. Since districts were phased into the program over this period, some individuals in these cohorts will have been under the program during their birth and childhood years and others will not. From *Column 7*, we see that eradication era cohorts in malarious areas experienced a 4 percentage point gain in literacy and a 6.7 percentage point gain in primary school completion. We do not see significant gains for the eradication era cohorts in potential epidemic areas.

Finally, we test whether our results are robust to the way we have classified districts into malaria prevalence groups based on the 1948 endemicity map. *Column 2* of *Table 8* restricts the sample to "unambiguous" districts, as defined in the data section. The effects in this restricted sample are actually slightly larger than the effects in the full sample; the literacy gain in malarious areas is 12.6 percentage points and the gain in potentially epidemic areas is 10.2 percentage points. The gain in primary school completion is 13.5 percentage points for malarious regions and 10.4 percentage points for potentially epidemic areas. All of these effects are strongly significant. *Column 3* restricts the sample to "unambiguous" and "majority" districts. The results are almost identical to those in the "unambiguous" sample alone.

5.2.2 District Convergence

A concern with the identification strategy is whether districts with high malaria prevalence followed the same trends over time relative to areas with low malaria prevalence. In particular, we might be concerned that educational attainment in potential epidemic and malarious areas would have converged toward non-malarious regions even in the absence of the eradication program. We perform two tests to check for possible convergence. First, we include post*region dummies in addition to district and year of birth fixed effects. A "region" is a sub-state geographic area defined by the NSS. There are 77 NSS regions and the median region contains 10 districts. This strategy allows us to identify the effects of malaria eradication using only *within-region* district variation in pre-eradication prevalence. Our results are robust to the inclusion of post*region controls. *Column 2* of *Table 9* shows that those born in potential epidemic areas or malarious areas experience a 6.8 percentage point increase in literacy relative to those born in non-malarious regions, and these effects are statistically significant at the 10% level. The point estimates for primary school completion are also very similar to those in the baseline specification, although only the effect for malarious regions is statistically significant. These results provide strong evidence that our findings are not driven by mere convergence across geographic areas.

We also perform a placebo regression to test whether individuals in potential epidemic and malarious areas experienced gains in educational attainment relative to those in non-malarious regions *prior* to the eradication program. We define the placebo eradication era as those birth cohorts born between 1933 and 1941. Placebo pre-eradication cohorts are the ten birth cohorts born between 1912 and 1932; placebo post-eradication cohorts are the ten birth cohorts born between 1942 and 52. *Column 3* presents the results. We do not find significant gains for those in potential epidemic or malarious areas relative to those in non-malarious areas for any educational outcome variable.[2]

5.3 Heterogeneity in Treatment Effects

We next test whether the effects of malaria eradication on educational outcomes vary by district or individual characteristics (*Table 10*). We have focused on the rural sample; *Column 2* shows the results of our baseline specification for the urban sample. The point estimates suggest that individuals in potentially epidemic or endemic districts had improvements in literacy and primary school completion of approximately 2.5 percentage points relative to those in non-malarious districts, effects that are about a quarter of the magnitude of those in the rural sample. The estimates for the urban sample are not statistically significant. As discussed, malaria was not nearly as serious a problem in urban India as rural India prior to the eradication era. These results suggest that the gains are not due to overall economic improvements in malarious areas that were unrelated to the eradication campaign, since such improvements would have most likely affected both rural and urban areas.

[2] We choose a specification with a placebo eradication era to allow the same "distance" between placebo treatment and control cohorts as in the true specification.

Returning to the rural sample, we compare the effects for females versus males. In malarious areas, males appear to benefit more from the eradication program than females. Males experience a 10.7 percentage point increase in literacy rate versus a 7.8 percentage point increase for females. The percentage point gains in primary school completion are 11.5 for males and 8.7 for females. The gender disparity in potential epidemic areas is more striking: we do not see significant gains in literacy or primary education completion for females in potential epidemic areas, while the gains for males are quite large (14 to 15 percentage points).

We also examine the effects for individuals that are members of a scheduled caste (SC) or scheduled tribe (ST) versus those that are not. Scheduled caste and scheduled tribes are groups that have been historically disadvantaged in India. We find that the effects for these groups are only slightly smaller than those for non-scheduled caste and tribe.

6 Conclusion

In 1953, the government of India launched a country-wide malaria eradication program that used the new DDT spraying technology to control and eradicate malaria. The ten-year program was hugely successful, with the officially estimated number of malaria cases driven from 75 million per year (21.8% of total population) to a mere 0.1 million per year (0.02% of total population). The number of recorded deaths from malaria was also dramatically reduced from 800,000 per year to less than 100 per year. In this paper, we combine individual-level data from the National Sample Survey (NSS) with district-level data on measures of malaria prevalence. Using the heterogeneity in indigenous malaria rates and the exogenous implementation of a malaria eradication program, we find that malaria eradication leads to an increase in literacy rate and years of schooling.

Overall, we find support for the belief that improvements in health and in the disease environment can have a causal effect on human capital accumulation. This may be an important channel for affecting growth. In future work, we plan to examine other economic outcomes.

.

References

[1] Acemoglu, Daron and Simon Johnson (2005) "Disease and Development: The Effect of Life Expectancy on Economic Growth," MIT mimeo.

[2] Alleyne, George and Daniel Cohen (2002) "The Report of Working Group I of the Commission on Macroeconomics and Health," WHO Commission on Macroeconomics and Health, April.

[3] Almond, Douglas (2005) "Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenze Exposure in the Post-1940 U.S. Population," Columbia University mimeo.

[4] Barreca, Alan (2007) "The Long-Term Economic Impact of In Utero and Postnatal Exposure to Malaria," UC David mimeo.

[5] Bleakley, Hoyt (2007) "Malaria in the Americas: A Retrospective Analysis of Childhood Exposure," University of Chicago mimeo.

[6] Bloom, David and David Canning (2005) "Health and Economic Growth: Reconciling the Micro and Macro Evidence," working paper.

[7] Bloom, David and Jeffrey Sachs (1998) "Geography, Demography, and Economic Growth in Africa," Brookings Papers on Economic Activity, 1998:2, 207-295.

[8] Davis, Kingsley (1956) "The Amazing Decline of Mortality in Underdeveloped Areas," American Economic Review, 46:2, May, 305-318.

[9] Desowitz, Robert (1991) *The Malaria Capers: More Tales of Parasites and People, Research and Reality,* New York: W.W. Norton.

[10] Duffy, P.E. and R.S. Desowitz (2001) "Pregnancy Malaria Throughout History: Dangerous Labors," in *Malaria in Pregnancy: Deadly Parasite, Susceptible Host*, ed. by P.E. Duffy, and M. Fried. Taylor and Francis.

[11] Gallup, John Luke and Jeffrey Sachs (2001) "The Economic Burden of Malaria," Am. J. Trop. Med. Hyg., 64, 85-96.

[12] Hoff, Brent and Carter Smith III (2000) *Mapping Epidemics: A Historical Atlas of Disease*, Franklin Watts, New York.

[13] Hong, Sok Chul (2007) "A Longitudinal Analysis of the Burden of Malaria on Health and Economic Productivity: The American Case," University of Chicago mimeo.

[14] Kiple, Kenneth (1993) *The Cambridge World History of Human Disease*, Cambridge University Press.

[15] Lucas, Adrienne (2005) "Economic Effects of Malaria Eradication: Evidence from the Malarial Periphery," Brown University mimeo.

[16] Lucas, Adrienne and David Weil (2005) "Social and Economic Effects of Malaria Eradication: Evidence from Sri Lanka," Brown University mimeo.

[17] McCormick, M.C., J. Brooks-Gunn, K. Workman-Daniels, J. Turner, and G.J. Peckman (1992) "The Health and Development Status of Very Low-Birth-Weight Children at School Age," Journal of the American Medical Association, 267:16.

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

[19] Moffitt, Robert (2004) "The Role of Randomized Field Trials in Social Science Research: A Perspective from Evaluations of Reforms of Social Welfare Programs," American Behavioral Scientist, 47:5, 506-540.

[20] NMEP (1996). *Epidemiology and Control of Malaria in India.* National Malaria Eradication Program.

[21] NMEP (1986). *Malaria and its Control in India, Vol. 1.* National Malaria Eradication Program.

[22] World Health Organization (2001) *Macroeconomics and Health: Investing in Health for Economic Development.*

[23] World Health Organization (1967). *The World Health Organization: Twenty Years in South-East Asia, 1948-1967.* New Delhi: World Health Organization, Regional Office for South-East Asia.

FIGURE 1.1: NMCP 1953-54

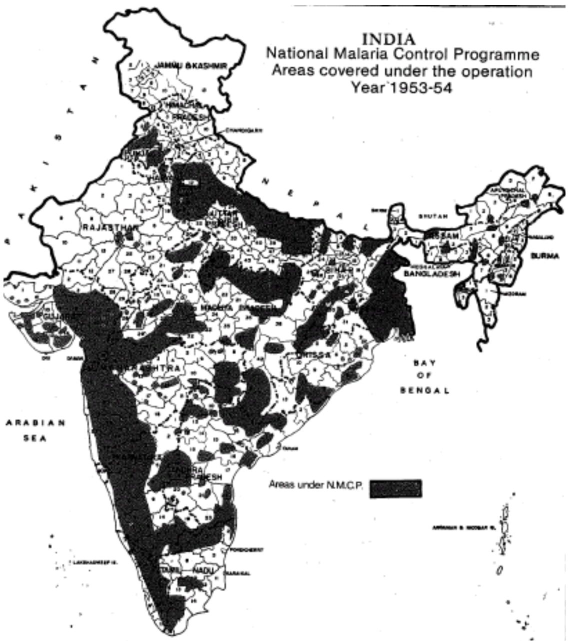

FIGURE 1.2: NMCP 1954-55

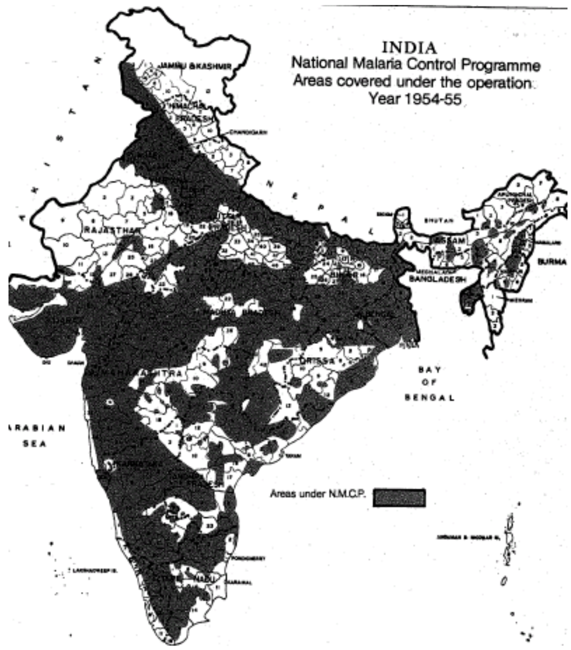

FIGURE 1.3: NMCP 1956-57

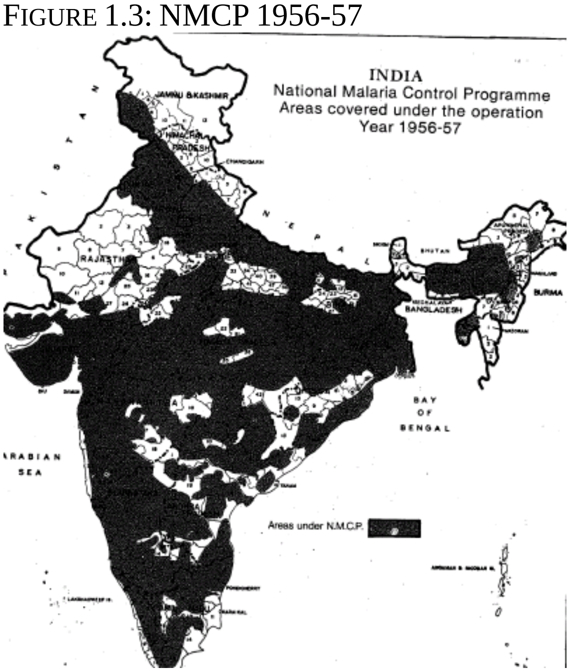

FIGURE 1.4: NMEP 1959-61

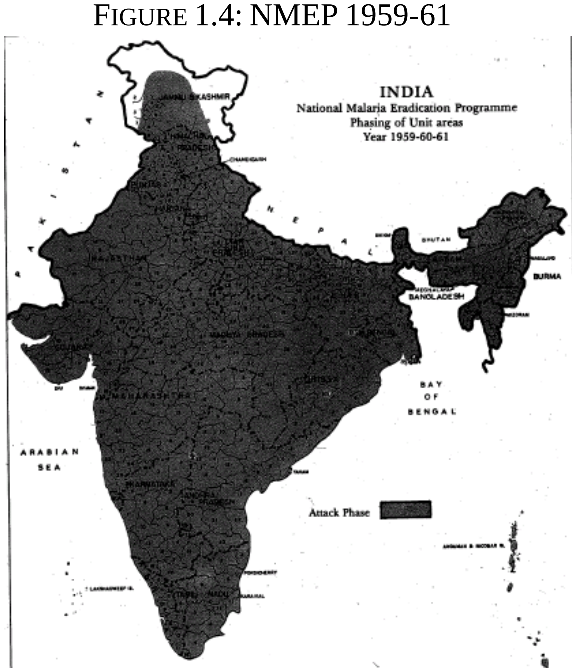

FIGURE 2: DECLINES IN MALARIA DEATHS FROM VITAL STATISTICS DATA

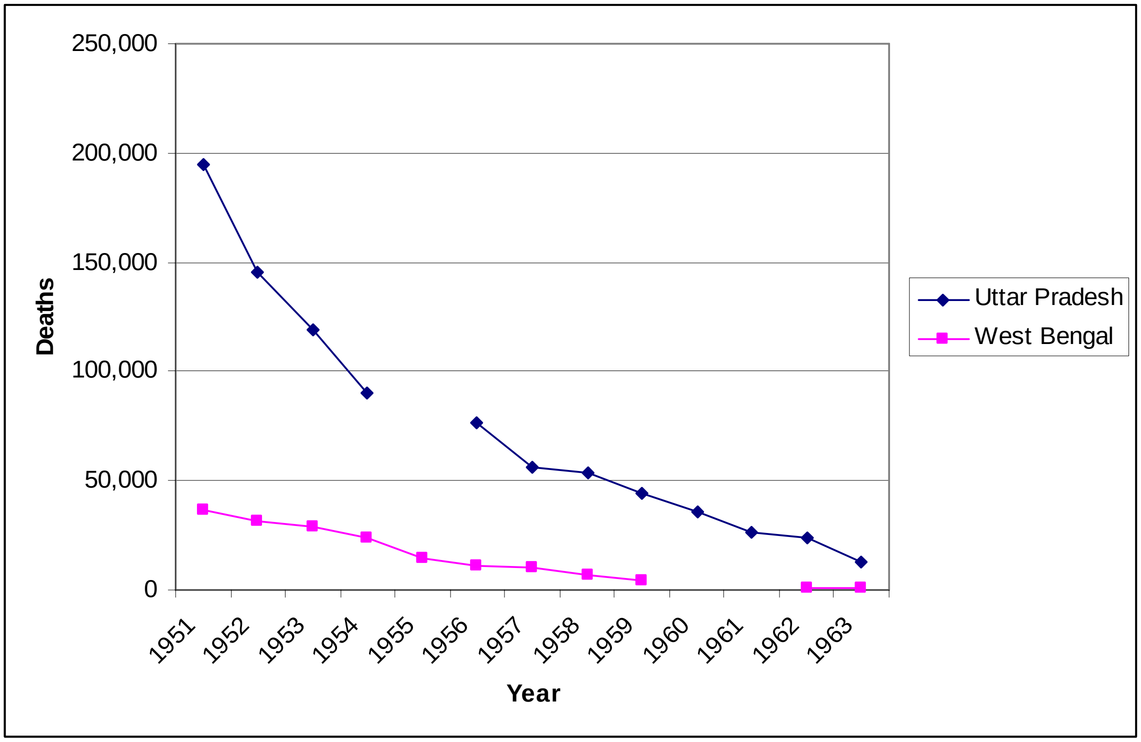

FIGURE 3: MALARIA ENDEMICITY MAP

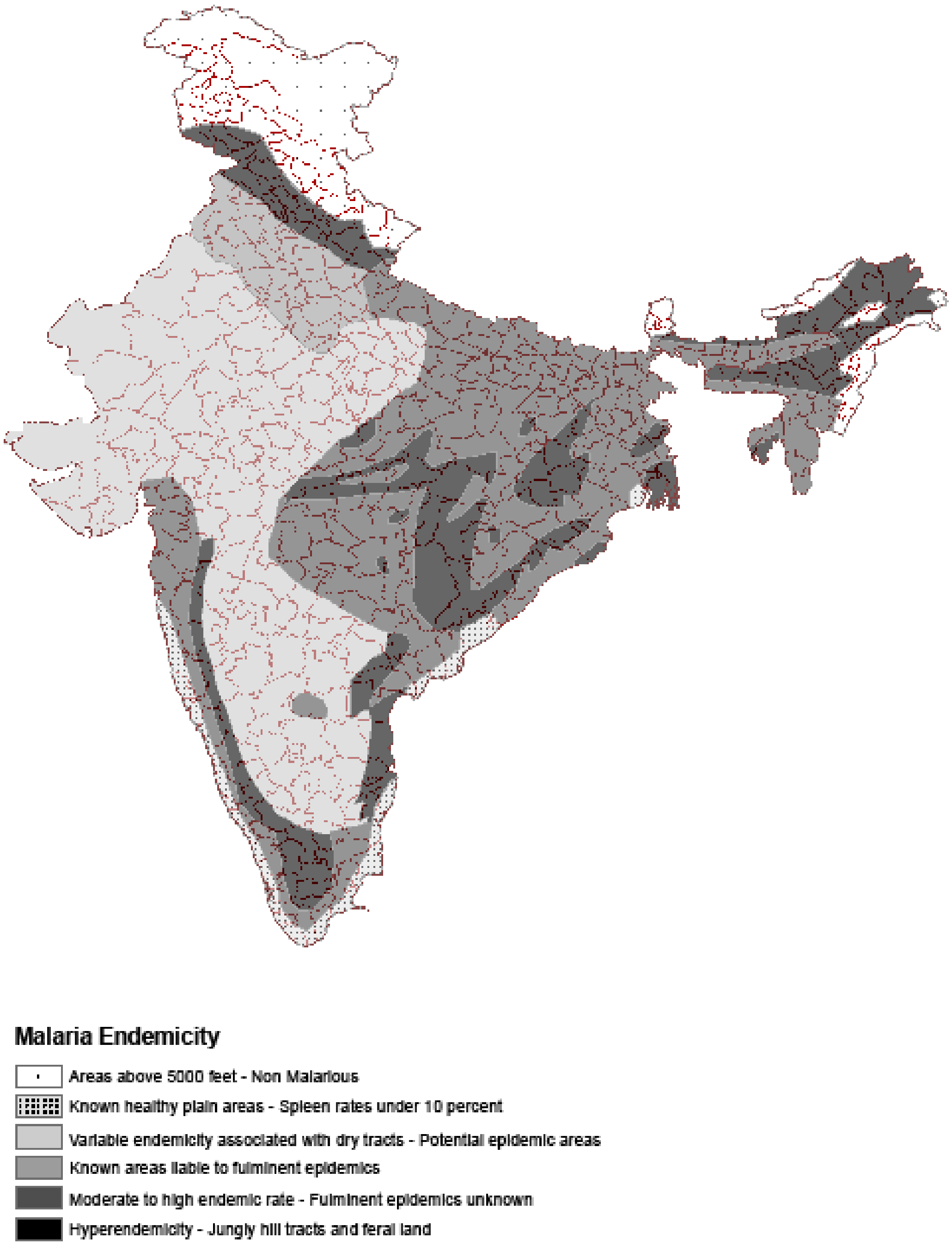

TABLE 1
Expansion of the National Malaria Control Program

	Number of Units Established	Number of Units Established (Cumulative)	Population Protected (in millions)	Population Protected (Cumulative, in millions)
1953-54	84.00	84.00	49.50	49.50
1954-55	26.75	110.75	30.40	79.90
1955-56	23.00	113.75	32.10	112.00
1956-57	35.50	169.25	32.50	144.50
1957-58	23.25	192.50	21.07	165.57

Source: NMCP (1986)

TABLE 2
Malaria Prevalence over Time

	Population (in millions)	Malaria cases (in millions)	Percentage of population with malaria each year	Number of deaths
1947	344	75.0	21.8	800,000
1965	466	0.1	0.02	0
1976	565	6.5	1.1	59
1984	710	2.2	0.3	247
1994	861	2.5	0.3	1122

Source: NMCP (1986), NMCP (1996)

TABLE 3
Malaria Prevalence Indicators during the NMCP

	Child Spleen Rate	Child Parasite Rate	Infant Parasite Rate
1953-54	15.7	3.9	1.6
1954-55	12.4	4.2	2.0
1955-56	7.7	1.8	0.7
1956-57	6.0	1.1	0.5
1957-58	4.2	0.8	0.6
1958-59	3.2	0.5	0.2
1959-60	1.4	0.16	0.14

Source: NMCP (1986)

TABLE 4

Summary Statistics for 43rd National Sample Survey (NSS)

	Mean	Standard Deviation
% Rural	0.673	(0.47)
% Female	0.404	(0.49)
% Married	0.530	(0.50)
Age	27.983	(17.87)
Household size	6.448	(3.19)
% Scheduled caste	0.140	(0.35)
% Scheduled tribe	0.117	(0.32)
% Hindu	0.770	(0.42)
% Muslim	0.127	(0.33)
% Christian	0.053	(0.22)
Number of states	31	
Number of regions	77	
Number of districts	431	
Number of households	291,648	
Number of observations	1,005,863	

Source: NSS (1987)

TABLE 5.1
Summary Statistics for Pre-eradication Cohort 1912-1952

Map	Literate	Primary	Middle	Secondary	College	Wage in cash	Wage in kind	Total wage
Non-malarious regions	0.363	0.247	0.136	0.066	0.020	4.589	0.420	5.010
	(0.481)	(0.431)	(0.343)	(0.248)	(0.139)	(33.823)	(7.063)	(34.659)
	8,813	8,813	8,813	8,813	8,813	3,665	3,665	3,665
Potential epidemic areas	0.268	0.163	0.089	0.045	0.008	5.156	4.439	9.595
	(0.443)	(0.370)	(0.285)	(0.208)	(0.091)	(33.375)	(499.091)	(500.226)
	43,649	43,649	43,649	43,649	43,649	19,623	19,623	19,623
Malarious regions	0.331	0.205	0.106	0.053	0.011	10.409	1.424	11.833
	(0.471)	(0.403)	(0.308)	(0.224)	(0.106)	(393.268)	(185.040)	(434.674)
	163,520	163,520	163,520	163,520	163,520	75,938	75,938	75,938
Total	0.317	0.196	0.102	0.051	0.011	9.159	2.086	11.245
	(0.465)	(0.397)	(0.303)	(0.221)	(0.103)	(343.995)	(286.132)	(447.476)
	215,982	215,982	215,982	215,982	215,982	99,226	99,226	99,226

Note: The first number is the mean, the second is the standard deviation, and the third is the number of observations. Summary statistics are for rural sample only.

TABLE 5.2
Summary Statistics for Post-eradication Cohort 1962-1972

Map	Literate	Primary	Middle	Secondary	College	Wage in cash	Wage in kind	Total wage
Non-malarious regions	0.497	0.418	0.240	0.164	0.026	1.534	0.401	1.936
	(0.491)	(0.500)	(0.474)	(0.370)	(0.158)	(13.727)	(5.237)	(15.319)
	6,438	6,438	6,438	6,438	6,438	2,260	2,260	2,260
Potential epidemic areas	0.456	0.374	0.230	0.098	0.012	6.128	2.775	8.904
	(0.498)	(0.484)	(0.421)	(0.297)	(0.109)	(354.429)	(316.979)	(475.633)
	32,523	32,523	32,523	32,523	32,523	13,570	13,570	13,570
Malarious regions	0.505	0.400	0.252	0.107	0.014	3.251	4.123	7.374
	(0.500)	(0.490)	(0.434)	(0.309)	(0.116)	(62.907)	(465.468)	(479.171)
	111,858	111,858	111,858	111,858	111,858	45,259	45,259	45,259
Total	0.317	0.196	0.102	0.051	0.011	9.159	2.086	11.245
	(0.465)	(0.397)	(0.303)	(0.221)	(0.103)	(343.995)	(286.132)	(447.476)
	215,982	215,982	215,982	215,982	215,982	99,226	99,226	99,226

Note: The first number is the mean, the second is the standard deviation, and the third is the number of observations. Summary statistics are for rural sample only.

TABLE 6

Baseline Results: DD Regression Using 1948 Malaria Endemicity Map

	(1) Literate	(2) Primary	(3) Middle	(4) Secondary	(5) College	(6) Wage in cash	(7) Wage in kind	(8) Total wage
Post * Malarious	0.093	0.101	0.052	0.041	0.009	4.094	-4.993	-0.899
	(0.036)***	(0.043)**	(0.028)*	(0.022)*	(0.007)	(2.510)	(3.701)	(4.501)
Post * Potential Epidemic	0.085	0.090	0.049	0.040	0.007	-1.325	2.383	1.057
	(0.036)**	(0.043)**	(0.028)*	(0.023)*	(0.007)	(1.258)	(2.838)	(3.115)
Female	-0.306	-0.226	-0.145	-0.076	-0.015	8.077	8.385	16.463
	(0.005)***	(0.004)***	(0.003)***	(0.002)***	(0.001)***	(4.797)*	(3.778)**	(6.117)***
Scheduled Caste	-0.181	-0.154	-0.103	-0.060	-0.013	-3.437	-1.105	-4.542
	(0.007)***	(0.006)***	(0.004)***	(0.003)***	(0.001)***	(1.976)*	(2.518)	(3.214)
Scheduled Tribe	-0.227	-0.180	-0.114	-0.064	-0.015	-4.963	2.284	-2.679
	(0.010)***	(0.009)***	(0.006)***	(0.004)***	(0.001)***	(1.549)***	(8.984)	(9.119)
Hinduism	0.002	-0.005	-0.014	-0.006	-0.001	-24.251	-0.750	-25.001
	(0.018)	(0.018)	(0.013)	(0.010)	(0.003)	(24.447)	(2.060)	(24.457)
Islam	-0.095	-0.122	-0.102	-0.055	-0.013	-28.218	-2.961	-31.179
	(0.022)***	(0.021)***	(0.015)***	(0.011)***	(0.003)***	(24.512)	(1.837)	(24.512)
Christianity	0.092	0.075	0.060	0.041	0.009	-30.484	-1.643	-32.128
	(0.028)***	(0.026)***	(0.020)***	(0.013)***	(0.004)**	(28.807)	(1.971)	(28.731)
Jainism	0.324	0.305	0.152	0.088	0.050	-28.426	290.652	262.226
	(0.061)***	(0.054)***	(0.039)***	(0.036)**	(0.023)**	(23.941)	(298.797)	(299.520)
Constant	0.197	0.091	0.158	0.070	0.022	8.772	-3.193	5.579
	(0.035)***	(0.040)**	(0.011)***	(0.008)***	(0.002)***	(3.405)**	(5.294)	(6.320)
Observations	366801	366801	304048	304048	304048	138593	138593	138593
R-squared	0.27	0.24	0.16	0.09	0.02	0.01	0.00	0.01

Note: Map categories 1 and 2 (omitted category) are defined as non-malarious regions. Map categories 3 and 4 are potential epidemic areas. Map categories 5 and 6 are malarious regions. Analysis focuses on rural sample only. Total sample excludes 1953-1961 cohorts born during the eradication era. Pre-eradication cohorts are defined as 1912-1952. Post-eradication cohorts are defined as 1962-1972 (age 15 or above) for columns (1) and (2), and are defined as 1962-67 (age 20 or above) for the other columns. Control variables include district and year of birth dummies, female, scheduled caste, scheduled tribe, and household religion. Robust standard errors clustered by district. ** significant at 5%; *** significant at 1%.

TABLE 7
Robustness Tests: Empirical Specification and Sample Selection

	(1) Baseline specification	(2) Exclude controls	(3) Exclude Jammu & Kashmir, Himachal Pradesh	(4) Pre-eradication = 1942-52	(5) Eradication era = 1950-1963	(6) Restrict to age 20+	(7) Include eradication-era cohort
Dependent Variable: Literate							
Post * Malarious	0.093	0.077	0.112	0.078	0.095	0.078	0.093
	(0.036)***	(0.036)**	(0.049)**	(0.036)**	(0.042)**	(0.032)**	(0.036)***
Post * Potential Epidemic	0.085	0.066	0.102	0.072	0.089	0.067	0.085
	(0.036)**	(0.036)*	(0.049)**	(0.036)**	(0.043)**	(0.033)**	(0.036)**
Eradication era * Malarious							0.040
							(0.023)*
Eradication era * Potential Epidemic							0.026
							(0.023)
Observations	366801	366801	331297	258373	209142	291556	458993
R-squared	0.27	0.15	0.27	0.27	0.27	0.27	0.27
Dependent Variable: Primary							
Post * Malarious	0.101	0.089	0.149	0.084	0.131	0.095	0.101
	(0.043)**	(0.043)**	(0.054)***	(0.037)**	(0.056)**	(0.041)**	(0.043)**
Post * Potential Epidemic	0.090	0.076	0.137	0.076	0.124	0.077	0.090
	(0.043)**	(0.043)*	(0.054)**	(0.037)**	(0.056)**	(0.041)*	(0.043)**
Eradication era * Malarious							0.067
							(0.035)*
Eradication era * Potential Epidemic							0.048
							(0.035)
Observations	366801	366801	331297	258373	209142	291556	458993
R-squared	0.24	0.15	0.24	0.23	0.24	0.23	0.24

Note: Baseline specification follows Table 6. Column (2) excludes the set of control variables. Column (3) excludes the states of Jammu & Kashmir and Himachal Pradesh. Column (4) narrows the pre-eradication window from 1912-52 to 1942-52. Column (5) expands the eradication window from 1953-1961 to 1950-1963. Column (6) restricts sample to age 20 or above (as opposed to age 15 or above in the baseline), so the post-eradication cohort is defined as 1962-67. Finally, column (7) includes all three periods (pre-eradication, eradication, and post-eradication). All regressions except Column (2) include district and year of birth dummies, as well as control variables including female, scheduled caste, scheduled tribe, and household religion. Robust standard errors clustered by district. ** significant at 5%; *** significant at 1%.

TABLE 8

Robustness Tests: Empirical Specification and Sample Selection

	(1) Baseline specification	(2) Restrict to “unambiguous” districts	(3) Restrict to “unambiguous” and “majority” districts	(4) Choosing the least malarious category possible	(5) Choosing the most malarious category possible
Dependent Variable: Literate					
Post * Malarious	0.093	0.126	0.116	0.072	0.125
	(0.036)***	(0.038)***	(0.038)***	(0.017)***	(0.041)***
Post * Potential Epidemic	0.085	0.102	0.102	0.047	0.116
	(0.036)**	(0.038)***	(0.038)***	(0.017)***	(0.041)***
Observations	366801	241682	294553	366801	366801
R-squared	0.27	0.26	0.27	0.27	0.27
Dependent Variable: Primary					
Post * Malarious	0.101	0.135	0.124	0.095	0.128
	(0.043)**	(0.047)***	(0.047)***	(0.018)***	(0.053)**
Post * Potential Epidemic	0.090	0.104	0.105	0.068	0.112
	(0.043)**	(0.047)**	(0.047)**	(0.018)***	(0.053)**
Observations	366801	241682	294553	366801	366801
R-squared	0.24	0.22	0.23	0.24	0.24

Note: Baseline specification follows Table 6. All regressions include district and year of birth dummies, as well as control variables including female, scheduled caste, scheduled tribe, and household religion. Robust standard errors clustered by district. ** significant at 5%; *** significant at 1%.

TABLE 9

Robustness Tests: District Convergence

	(1) Baseline specification	(2) Include regional trends	(3) Placebo test: pre=1912-32, post=1942-52	(4) Placebo test: pre=1912-42, post=1943-52
Dependent Variable: Literate				
Post * Malarious	0.093	0.068	0.060	0.024
	(0.036)***	(0.039)*	(0.038)	(0.026)
Post * Potential Epidemic	0.085	0.068	0.057	0.019
	(0.036)**	(0.040)*	(0.038)	(0.026)
Observations	366801	366550	169804	215982
R-squared	0.27	0.28	0.26	0.26
Dependent Variable: Primary				
Post * Malarious	0.101	0.105	0.051	0.043
	(0.043)**	(0.055)*	(0.043)	(0.032)
Post * Potential Epidemic	0.090	0.078	0.041	0.032
	(0.043)**	(0.055)	(0.043)	(0.033)
Observations	366801	366550	169804	215982
R-squared	0.24	0.25	0.20	0.19

Note: Baseline specification follows Table 6. Column (2) allows for differential regional trends by adding in post*region dummies. Columns (3) and (4) present results for placebo tests. All regressions include district and year of birth dummies, as well as control variables including female, scheduled caste, scheduled tribe, and household religion. Robust standard errors clustered by district. ** significant at 5%; *** significant at 1%.

TABLE 10
Heterogeneity in Treatment Effects

	(1) Rural	(2) Urban	(3) Rural Female	(4) Rural Male	(5) Rural SC / ST	(6) Rural non-SC & non-ST
Dependent Variable: Literate						
Post * Malarious	0.093	0.024	0.078	0.107	0.069	0.096
	(0.036)***	(0.049)	(0.038)**	(0.061)*	(0.036)*	(0.039)**
Post * Potential Epidemic	0.085	0.025	0.044	0.137	0.070	0.086
	(0.036)**	(0.049)	(0.038)	(0.061)**	(0.036)*	(0.039)**
Observations	366801	184282	146149	220401	110477	256324
R-squared	0.27	0.24	0.28	0.19	0.23	0.27
Dependent Variable: Primary						
Post * Malarious	0.101	0.025	0.087	0.115	0.077	0.099
	(0.043)**	(0.053)	(0.048)*	(0.062)*	(0.020)***	(0.051)*
Post * Potential Epidemic	0.090	0.020	0.047	0.150	0.073	0.088
	(0.043)**	(0.054)	(0.048)	(0.062)**	(0.021)***	(0.051)*
Observations	366801	184282	146149	220401	110477	256324
R-squared	0.24	0.23	0.24	0.19	0.19	0.24

Note: Baseline specification follows Table 6. All regressions include district and year of birth dummies, as well as control variables including female, scheduled caste, scheduled tribe, and household religion. Robust standard errors clustered by district. ** significant at 5%; *** significant at 1%.