# DISEASE AND DEVELOPMENT: EVIDENCE FROM HOOKWORM ERADICATION IN THE AMERICAN SOUTH*

## HOYT BLEAKLEY

This study evaluates the economic consequences of the successful eradication of hookworm disease from the American South, which started *circa* 1910. The Rockefeller Sanitary Commission (RSC) surveyed infection rates and found that 40 percent of school-aged children in the South were infected with hookworm. The RSC then sponsored treatment and education campaigns across the region. Follow-up studies indicate that this campaign substantially reduced hookworm disease almost immediately. Areas with higher levels of hookworm infection prior to the RSC experienced greater increases in school enrollment, attendance, and literacy after the intervention. No significant contemporaneous results are found for literacy or occupational shifts among adults, who had negligible prior infection rates. A long-term follow-up indicates a substantial gain in income that coincided with exposure to hookworm eradication. I also find evidence that the return to schooling increased with eradication.

## I. INTRODUCTION

The importance of the burden of tropical disease in impeding economic development has received considerable attention in recent years. The establishment and maintenance of an environment free of infectious disease is an important public good. The very nature of the transmission mechanism of such diseases implies a manifest externality. This might serve as a rationale for collective action to reduce the incidence of infectious disease. However, little is known about the long-term benefits of such actions, and therefore there is nothing to compare with the short-term costs.

Unfortunately, simple correlations of public health and economic outcomes are unlikely to measure the causal effect since

public health is endogenous. Indeed, it is likely a normal good: rich areas purchase more of it. To measure the contribution of a disease-free environment, we need to analyze plausibly exogenous improvements in public health. Targeted public-health interventions are a possible source of such variation.

The present study focuses on one specific intervention targeted toward hookworm disease in the American South. The hookworm-eradication campaign (*circa* 1910–1915) began soon after (i) the discovery that a variety of health problems among Southerners could be attributed to the disease and (ii) the donation by John D. Rockefeller of a substantial sum to the campaign. The Rockefeller Sanitary Commission (RSC) surveyed infection rates in the affected areas, and found that an average of 40 percent of school-aged children in the American South suffered from hookworm infection. The RSC then sponsored treatment dispensaries that traveled these areas providing deworming medications and educating local physicians and the public about prevention. Follow-up studies indicate that the campaign brought about a substantial immediate reduction in hookworm disease and, furthermore, that the seeds were sown for preventing its return.

The introduction of this treatment (broadly defined) combines with the cross-area differences in pretreatment infection rates to form the basis of my identification strategy. As the RSC surveys demonstrated, different areas of the country had distinct incidences of the hookworm disease. Areas with high infection rates had more to gain from the newly available treatments, whereas areas with little hookworm disease did not. This heterogeneity allows for a treatment/control strategy.

Moreover, the eradication campaign began—and was ultimately successful—because of critical innovations to knowledge. I argue that such innovations were not related to or somehow in anticipation of the future growth prospects of the affected areas, and therefore should not be thought of as endogenous in this context. For example, the discovery of the transmission mechanism for hookworm was made by a European doctor whose initial experimental evidence consisted of accidentally infecting himself while diagnosing a patient. At that time, hookworm infection in the American South was not even recognized as a problem.

Hookworm disease, while rarely fatal, has potentially severe chronic symptoms. The hookworm is a parasite that lodges itself in the victim's digestive system, and burrows into the intestinal

wall and tapping into the host's bloodstream. Listlessness, anemia, and stunting of growth are common symptoms among infected children. Because schoolwork is an energy-intensive activity for children, it is plausible that hookworm disease would depress the returns to human-capital investment.

After hookworm eradication, school enrollment, regular school attendance, and literacy increased markedly in counties that had previously suffered from high rates of hookworm infection. This is true in absolute terms as well as relative to comparison counties that had lower levels of hookworm infection. I find this result using either a two-period double difference or a multiperiod setup that allows for differential trends across areas. Furthermore, the conclusion is robust to controlling for a variety of other alternative hypotheses, including crop-specific shocks, demographic shifts, the near-simultaneous reduction in malaria, parental socioeconomic status, and certain policy changes. Estimates using indirect least squares imply that a child infected with hookworm had a 20 percent lower probability of school enrollment, although it is impossible to completely rule out that the intervention had effects through channels besides measured hookworm infection. Replicating this design using state-of-birthlevel variation in hookworm infection yields similar estimates for these variables, although the results for enrollment are imprecise.

Next, I present analogous results for adults as a specification check. *A priori* we would expect that adults would be substantially less affected by the hookworm-eradication campaign because adults were substantially less likely to have hookworm [RSC 1911; Smillie and Augustine 1925]. Moreover, human-capital investments not made in childhood due to hookworm would be water under the bridge once the disease environment improved. On the other hand, if the results for children were due to changes in income or migration patterns, we would see changes in adult outcomes as well. Instead, I find evidence that there was little contemporaneous impact on adults, measured along several important dimensions: literacy, labor-force participation, and occupation.

I also follow up on the cohorts that potentially benefited from hookworm eradication during childhood. Here I contrast individuals based on (i) the pre-eradication hookworm burden in their state of birth and (ii) their year of birth relative to the RSC. Cohorts more exposed to the eradication efforts went on to earn substantially higher incomes as adults. This pattern is seen using

data on wage and salary incomes from the 1940 Census. Again using indirect least squares (and subject to the same *caveat lector* as above). I estimate that being infected with hookworm through one's childhood led to a reduction in adult wages of approximately 40 percent. I also consider occupational proxies of income, which are defined over a broad range of Census years, and show that the shift in the hookworm-income relationship coincides with childhood exposure to the eradication campaign, rather than with some preexisting trend or autoregressive process. No statistically significant long-term effect of hookworm is found on the years of schooling (in accordance with the imprecise result for enrollment using state variation), but both literacy and returns to schooling increased with exposure to hookworm eradication.

The rest of this study is organized as follows. Section II describes the symptoms and history of the disease. Section II.D discusses in particular how the circumstances of the discovery of the hookworm problem in the South and the subsequent anti-hookworm campaign lend themselves to a strategy for identifying the effect of hookworm. Section III describes the data employed. The contemporaneous results using sequential cross sections are presented in Section IV and V. The long-term follow-up is found in Section VI. I conclude the study in Section VII.

## II. Hookworm and the Rockefeller Sanitary Commission

### II.A. *Hookworm Disease*

Hookworm is an intestinal parasite that lodges itself in the human intestine and absorbs nutrients from the victim's bloodstream. The symptoms of hookworm infection (or *uncinaria*) are lethargy and anemia. In rare cases, the anemia can become so severe as to cause death. The life cycle of the hookworm is dependent on unsanitary conditions. The nematodes lay their eggs in the intestine, but the larvae are passed out of the digestive system in feces. Hookworm is therefore transmitted through skin contact with infected fecal matter. The larvae then burrow their way in through the skin. The lifespan of a hookworm is much shorter than that of a human, and so continuous reinfection is required to generate any sustained worm load.

There are two angles for managing hookworm: treatment and prevention. The treatment consists of simply taking a deworming medicine. Preventative measures include limiting skin

contact with polluted soil (through the use of shoes, for example) and dealing with excrement in ways that minimize soil pollution in the first place (e.g., the use of sanitary latrines).

## II.B. The Eradication Campaign

The Rockefeller Sanitary Commission for the Eradication of Hookworm Disease was formed in 1910 with the donation of one million dollars by John D. Rockefeller. Some years before, an American doctor (Charles W. Stiles) had recognized hookworm symptoms in Southerners. Through intermediaries, Dr. Stiles had convinced Rockefeller that taking on hookworm was a good foray into large-scale charity. The Commission began by conducting surveys of hookworm-infection rates among children across the region. The RSC surveyed over 600 counties in the South and found hookworm infection to be over 40 percent among children.

Soon after, the treatment campaign began. First, the RSC sent teams of health-care workers to counties to administer and dispense deworming treatments free of charge. RSC dispensaries visited a large and mostly contiguous fraction of the South and the campaign treated over 400,000 individuals with deworming medication.[1] Second, the RSC sought to educate doctors, teachers, and the general public on how to recognize the symptoms of hookworm disease so that fewer cases would go untreated. Another part of this publicity campaign included education about the importance of hygiene, especially with regard to the use of sanitary privies. In this period, oftentimes even public buildings such as schools and churches did not have such hygienic facilities. Follow-up surveys conducted afterward showed a substantial decline in hookworm infection [RSC 1915]. Although the stated goal of eradication was not achieved, the hookworm-infection rate of the region did drop by more than half, and fewer extreme cases of the disease went unnoticed and untreated.

Because the deworming treatments are short-term solutions, eradication requires (a) sustained monitoring (and treatment as needed) and (b) a reduction in the probability of reinfection. Follow-up efforts by private and governmental actors likely played a key role in consolidating the gains from the RSC and continuing the progress toward complete eradication.[2] State gov-

---

1. Thymol, taken orally, was the recommended treatment of the time.
2. An interesting episode for comparison comes from Puerto Rico. Around the same time as the RSC, a commission from the U.S. Army–sponsored treatment

ernments ramped up their funding of anti-hookworm campaigns as the RSC was winding down. Local and state governments eventually took over some of its activities. The successor to the RSC, the Rockefeller Foundation's International Health Board (IHB), continued to be involved at a lower level of funding. The IHB sponsored a handful of demonstration projects of the "intensive method," which combined the deworming treatments and publicity campaigns of the RSC with technical assistance in building latrines at homes and public buildings. The state boards of health largely adopted this method and applied it to a degree throughout their jurisdictions. Harder to measure, but of considerable importance, the hookworm problem had entered into the public consciousness.

## II.C. Testimonials Following the Campaign

Anecdotal evidence suggests that the RSC had an impact on human capital. Periodically, educators would write the Commission thanking it for its efforts and describing the improvements following hookworm treatment. The following letter is from the school board of Varnado, Louisiana. [RSC 1912].

> As a result of your treatment for hookworm in our school, we find that children who were ranking fifth and sixth in their classes now rank second and third. Their lessons are not so hard for them: they pay better attention in class and they have more energy. [ . . . ] In short, we have here in our school-rooms today about 120 bright, rosy-faced children, whereas had you not been sent here to treat them we would have had that many pale-faced, stupid children.

Farmer [1970] relates the following testimonials from the same period:

> Teachers, school officials, and editors continued to be amazed at the difference in children after treatment for hookworm disease. A. J. Caldwell, Principal of Hammond High School in Louisiana, wrote that there was a decided improvement in the students in his school. One girl, who was in the

---

and education campaigns throughout that Caribbean island. Large gains against hookworm were realized immediately after the campaign. Unfortunately, the colonial government provided very little follow-up support, and these gains had almost completely disappeared a decade later. Moreover, recent work in Kenya by Kremer and Miguel [2004] suggests that the initial impulse provided by short-term injections of medication and publicity may have little long-term benefits. The present study is not a guide to how to design a deworming program, and thus I do not take a stand on the relative merits of medications versus publicity versus sanitation. Rather, I take the reduction in hookworm as given and evaluate the socioeconomic consequences.

fifth grade and did not attend school regularly because she was so pale and weak, started regaining her color and strength after treatment and finished the school term at the top of her class. C. C. Wright, Superintendent of Schools in Wilkes County, North Carolina, was an ardent supporter of the eradication program after examination of the pupils in his district revealed over 50 percent infection. Treatment cured the majority of these cases and the quality of performance in the county schools was raised considerably.

Typical of school officials' attitude was that of W. H. Smith, State Supervisor of Rural Schools in Mississippi, who was thoroughly convinced that the economic prosperity of the people and the progress of educational development of the state depended largely on the successful eradication of the hookworm. The mental and physical growth of hundreds of children was evident. Smith asked for expansion of the program so that the thousands of children who were still suffering from mental and physical retardation might be saved.

And a report [RSC 1915] describes a Tidewater-Virginia community's experience of

how the treatment of these children had transformed the school. Children who were listless and dull are now active and alert; children who could not study a year ago are not only studying now, but are finding joy in learning. These children were born of anemic parents; were themselves infected in infancy; for the first time in their lives their cheeks show the glow of health. With this has come a new light to the eye, a new spring to the step, a new outlook on life. All this shows itself in a new spirit in the school. [ . . . ] Some of the 15 children who had never attended school, having been treated, have come in during the year. Others have declared their intention to enter in the fall.

## II.D. Identification Strategy

The first factor for identifying the effect of the hookworm-eradication campaign is that different areas of the South had distinct incidences of the disease. Hookworm larvae were better equipped to survive in areas with sandy soil and a warm climate. Broadly, this meant that the residents of the coastal plain of the South were much more vulnerable to infection than were those from the piedmont or mountain regions. Populations in areas with high (preexisting) infection rates were in a position to benefit from the newly available treatments, whereas areas with low prevalence were not. This heterogeneity allows for a treatment-control strategy.

Second, the initiation of the campaign by the RSC was largely a function of factors external to the Southern states.[3] The

---

3. The historical presentation in this section draws heavily on the work of Ettling [1981].

eradication campaign was made possible by critical innovations to knowledge: understanding how the disease worked and more importantly recognizing its presence. This contrasts with explanations that might have troublesome endogeneity problems, such as changes in government spending or positive income shocks in the infected areas. But even with the knowledge of the hookworm problem, there would have been formidable obstacles to taking action. The public-health infrastructure of this period was extremely limited. Rockefeller's donation was an important precondition for attacking the problem.

Thirdly, the anti-hookworm campaign achieved considerable progress against the disease in less than a decade. This is a sudden change on historical time scales. Moreover, I examine outcomes over a fifty-year time span, which is unquestionably long relative to the five-year RSC intervention.

These factors combine to form the central variable in the present study:

(Pretreatment Infection Rate)$_j$ × (Indicator for Posttreatment)$_t$.

More compactly, call this variable ($H_j^{\mathrm{pre}} \times \mathrm{Post}_t$), where $j$ indexes the geographic area and $t$ indicates the year. The variable $H_j^{\mathrm{pre}}$ denotes the level of hookworm infection among school-aged children in area $j$ at the time of the RSC's initial survey, and $\mathrm{Post}_t$ is a dummy variable indicating whether year $t$ is later than the active years of the RSC campaign (1910–1915).

I compare the evolution of outcomes (e.g., investment in human capital) across counties with distinct hookworm-infection rates, in order to assess the contribution of the eradication campaign to the observed changes. Estimating (1) measures the reduced-form differences by pre-eradication hookworm for some outcome $Y_{ijt}$ for person $i$ in area $j$ at time $t$.[4]

$$(1) \qquad Y_{ijt} = \beta(H_j^{\mathrm{pre}} \times \mathrm{Post}_t) + \delta_t + \delta_j + X_{ijt}\Gamma + \mathcal{E}_{ijt}$$

in which $Y_{ijt}$ is the outcome of interest, the $\delta_t$ are time dummies, the $\delta_j$ are geographic fixed effects, and $X_{ijt}$ is some vector of individual-level controls.[5]

---

4. All of the estimates of this equation below are calculated using ordinary least squares (OLS) regressions.

5. The model is derived as follows. For individual $i$, in area $j$, in year $t$, we start with an individual-level model with individual infection data and linear effects of hookworm:

$$Y_{ijt} = \alpha H_{ijt} + \delta_j + \delta_t + X_{ijt}\Gamma + \overline{\mathcal{E}}_{ijt}$$

How realistic is the assumption that areas with high infection rates benefited more from the eradication campaign? Resurveys found a decrease in hookworm infection of 30 percentage points across the infected areas of the South. Such a dramatic drop in the region's average infection rate, barring a drastic reversal in the pattern of hookworm incidence across the region, would have had the supposed effect of reducing infection rates *more* in highly infected areas than in areas with moderate infection rates. Figure I presents data on this issue.[6] The basic assumption of this section that areas where hookworm was highly endemic saw a greater drop in infection than areas with low infection rates is born out across states and across counties.

## II.E. Related Studies

Several pieces of contemporaneous evidence also complement the results from the present study. Summarizing evidence from randomized trials in developing countries, Dickson et al. [2000] find mixed evidence of the effect of hookworm infection on schooling, whereas Miguel and Kremer [2004] estimate the impact to be strong and positive using an experiment in Kenya. Miguel and Kremer argue that infection spillovers contaminated the earlier mixed results. Specifically, previous studies often randomized *within* schools, but fail to deal with the reinfection problem. As a result, they argue, follow-up surveys often found limited effects; no increase in school attendance is observed because there is little persistent difference in infection rates between control and treat-

---

where $H_{ijt}$ is a dummy for being infected. Individual infection data are not available, so the hookworm infection rate $H_{ijt}$ is replaced with its ecological (i.e., aggregate) counterpart:

$$Y_{ijt} = \tilde{\alpha} H_{jt} + \delta_j + \delta_t + X_{ijt}\Gamma + \overline{\varepsilon}_{ijt}$$

(This equation can equally be run in aggregate form entirely, and, when estimated, it gives very similar results to those in the present manuscript.) For the instrument ($H_j^{\text{pre}} \times \text{Post}_t$), the reduced form of this system is (1). Alternatively, one could have written the individual-level model with separate terms for individual and aggregate infection variables, the latter reflecting some spillover from peer infection to own human capital. But both of these effects would be subsumed into the $\tilde{\alpha}$ coefficient on the ecological infection rate, and it is this composite coefficient that I seek to measure in the present study. I have also experimented with nonlinear specifications, but no robust pattern emerges for the curvature of the response to hookworm. I report linear specifications below.

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

$$H_{jt} = \gamma(H_j^{\text{pre}} \times \text{Post}_t) + \delta_j + \delta_t + \eta_{jt}$$

This equation can be written in first-differenced form and evaluated in the post-RSC period:

$$\Delta H_j^{\text{post}} = \gamma H_j^{\text{pre}} + \text{constant} + \nu_{jt}.$$

an equation that relates the observable variables graphed in Figure I.
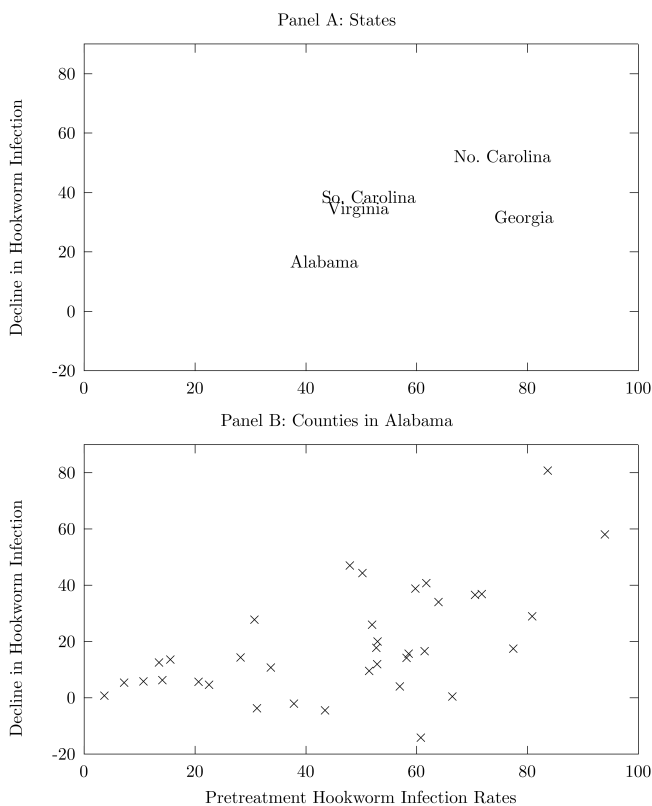
Panel A: States

Panel B: Counties in Alabama

FIGURE I

Highly Infected Areas Saw Greater Declines in Hookworm

The *y* axis displays the decrease in hookworm infection postintervention, as measured by follow-up surveys. The *x* axis is the pretreatment hookworm infection rate, as measured by the Rockefeller Sanitary Commission. Panel A displays data at the state level, as reported by Jacocks [1924]. Panel B contains data from counties in Alabama, as reported by Havens and Castles [1930]. Both resurveys are from the early 1920s. The average number of children examined per country exceeds 450 in both studies.

ment groups. (Philipson [2000] also discusses this evaluation issue in a general context.) Small-scale interventions that do not manage reinfection are therefore less likely to succeed. The RSC intervention, on the other hand, was of such a scale that it brought about large reductions in hookworm disease in entire areas, and these gains were further consolidated through improvements in sanitation. In the context of economic development, it is precisely such a large and persistent reduction in disease burden that we would wish to consider.

There are several other recent studies that consider the early-twentieth-century reduction in tropical diseases in the American South. While childhood effects are the focus of this study, Brinkley [1994] examines the role hookworm played in agricultural productivity. He finds a negative conditional correlation between hookworm infection and agricultural income per capita, although he does not specifically use the RSC intervention to identify this relationship. Bleakley [2002a] examines the interaction between malaria and hookworm. Bleakley and Lange [2004] consider the hookworm-related increase in returns to schooling in a quantity-quality model, and examine the fertility behavior of households in response to hookworm eradication. Additionally, an earlier version of the present study was found in the first chapter of Bleakley [2002b], and those results were partially summarized by Bleakley [2003]. The latter, summary piece discussed results for schooling and income, but did not treat literacy or regular school attendance nor did it consider whether hookworm-related changes in adult income were an artifact of some alternative time-series process, nor did it report results considering possible omitted variables.

## III. Data and Descriptive Statistics

This study links aggregate data on hookworm infection with individual socioeconomic data. Table I contains summary statistics of various outcomes. Because county boundaries change, I use aggregated county groupings, or "State Economic Areas" (SEA) as the geographic unit.

The hookworm-infection rates were computed by the Rockefeller Sanitary Commission for more than five hundred and fifty counties across the South. The data collection took place between 1910 and 1914 (at a single point in time for each county), and the summary statistics were constructed from samples of school-aged children in each county.[7] The RSC surveys measured an unweighted average infection rate across SEAs of 32 percent.

---

7. The infection rates were based on microscopic examination of stool samples. (Several microscopists were generally part of the survey and dispensary teams.) The following quote is from the Second Annual Report of the RSC [1911]:

> The survey is made by counties: it is based on a microscopic examination of foecal specimens from at least 200 children between the ages of six and eighteen taken at random—that is, without reference to clinical symptoms—from rural districts distributed over the county.

TABLE I
SUMMARY STATISTICS

| | Whole Sample | By Hookworm Infection | | Source |
|---|---|---|---|---|
| | | >40% | <40% | |
| Hookworm-Infection Rate | 0.320 | 0.554 | 0.164 | RSC Annual |
| | (0.230) | (0.137) | (0.117) | Reports |
| Individuals Treated At | 0.206 | 0.342 | 0.109 | RSC Annual |
| Least Once by the | (0.205) | (0.199) | (0.147) | Reports |
| RSC, Per School-Age | | | | |
| Child | | | | |
| School Enrollment, 1910 | 0.721 | 0.711 | 0.729 | IPUMS; author's |
| | (0.104) | (0.099) | (0.108) | calculations |
| Change in School | 0.089 | 0.103 | 0.078 | IPUMS; author's |
| Enrollment, 1910–1920 | (0.080) | (0.090) | (0.072) | calculations |
| Full-time School | 0.517 | 0.469 | 0.551 | IPUMS; author's |
| Attendance, 1910 | (0.140) | (0.123) | (0.141) | calculations |
| Change in Full-time | 0.203 | 0.246 | 0.172 | IPUMS; author's |
| School Attendance, | (0.097) | (0.093) | (0.089) | calculations |
| 1910–1920 | | | | |
| Literacy, 1910 | 0.853 | 0.824 | 0.875 | IPUMS; author's |
| | (0.104) | (0.101) | (0.102) | calculations |
| Change in Literacy, | 0.060 | 0.081 | 0.045 | IPUMS; author's |
| 1910–1920 | (0.067) | (0.075) | (0.057) | calculations |
| Population Black, 1910 | 0.357 | 0.41 | 0.318 | IPUMS; author's |
| | (0.221) | (0.208) | (0.223) | calculations |
| Fraction Population | 0.174 | 0.167 | 0.180 | ICPSR |
| Urban, 1910 | (0.200) | (0.214) | (0.190) | (1984) |
| School Term, in Months, | 5.251 | 5.055 | 5.391 | State annual |
| c. 1910 | (1.066) | (1.042) | (1.068) | reports |
| School per Square Mile, | 0.195 | 0.142 | 0.233 | State annual |
| c. 1910 | (0.358) | (0.053) | (0.465) | reports; ICPSR |
| Value of School Property, | 5.518 | 4.699 | 6.104 | State annual |
| per Pupil, Current | (4.037) | (3.159) | (4.496) | reports |
| Dollars, c. 1910 | | | | |
| Teacher-to-School Ratio, | 1.336 | 1.397 | 1.293 | State annual |
| c. 1910 | (0.545) | (0.505) | (0.572) | reports |
| Sample Size | 115 | 48 | 67 | n/a |

Variable means displayed to the right of variable name. Standard deviations displayed in parentheses below the mean. Sample selection for the IPUMS data: native-born whites and blacks, in the RSC-surveyed geographic units, for the indicated years. The school enrollment and attendance data are constructed from children aged 8–16; literacy data are for children 10–16, and the RSC reported infection rates for children aged 8–16. See the Appendix for further information for sources and variable construction.

The number of individuals treated at least once by the RSC (scaled by 1910 SEA youth population) are reported in Table 1. The second and third columns display the means by subsamples that are separated based on the intensity of their hookworm problem. Because of the policy of treating any infected person who presented himself at a Commission dispensary, the RSC directed more resources toward the areas with greater hookworm infection. These number indicate about 64 percent of the infected population was given deworming treatments.[8]

The micro-level data come from the *Integrated Public Use Micro Sample* (IPUMS), a project harmonizing the coding of historical Census microdata [Ruggles and Sobek 1997]. The RSC's activities took place from 1910–1915, therefore the core component of the data come from the decennial censuses that bracket the intervention. The sensitivity analysis uses census micro data from 1900–1950, and the long-term followup comprises census samples from 1880–1990.

Three binary indicators of human capital are used in the present study: school enrollment, regular or "full-time" school attendance, and literacy. The enrollment variable measures whether the child had gone to school for at least one day in the months preceding the Census.[9] I proxy for regular or "full-time" school attendance by combining the enrollment variable with occupational information. Children were coded as attending school regularly if both enrolled in school and not reporting an occupation. The literacy variable indicates whether the child can read and write.

The data show faster increases from 1910 to 1920 in the enrollment, attendance, and literacy rates in areas with high hookworm infection, coupled with lower average levels of these measures in 1910. The fact that this period coincides with the hookworm-eradication campaign is *prima facie* evidence that the

8. Similarly, a regression with these measures at the SEA level *(N = 113 and $R^2$ = .495)* yields the following estimates:
$$Tr_j = .619 H_j^{pre} + .003 + \varepsilon_j$$
$$\phantom{Tr_j = }(.064) \phantom{H_j^{pre} +} (.017)$$
where $Tr_j$ is the number of individuals treated at least once by the RSC, divided by the school-aged population (ages six to seventeen, inclusive) in SEA *j*. This indicates that, on the margin, about 62 percent of sufferers were treated.

9. The underlying Census question used the word "attendance" rather than "enrollment," but I call the variable "enrollment" nonetheless. The rather low standard of attending at least one day maps more closely onto enrollment, as the word is used in the contemporary literature.

increase in school attendance was related to the reduction in
hookworm disease.

Areas with greater hookworm burdens were different along
other margins as well. For one, they were more rural, and had
higher proportions of black residents. Additionally, the hook-
worm-infested areas also had shorter school terms and a lower
capital stock invested in primary education. There were also more
teachers per school, in part because of prevalence of one-room
common schools. These variables and others will be important
controls in the sensitivity analysis below.[10]

## IV. Contemporaneous Effects on Children

### IV.A. Main Results

In this subsection, I conduct regression analyses of changes
in literacy, school enrollment, and school attendance between the
1910 and 1920 Censuses by estimating (1) above. Using the
two-period comparison, I find a substantial increase in school
enrollment among children living in areas that had high levels of
hookworm infection in 1910. This is true in absolute terms and
also relative to areas with lower levels of infection. Specifically,
the coefficient on ($H_j^{\mathrm{pre}} \times \mathrm{Post}_t$) implies that a county with a 1910
infection rate of 50 percent would experience an increase in school
enrollment of 3–5 percentage points, relative to a county with no
infection problem. In 1910, the mean of school enrollment in the
sample was 0.78 and the standard deviation across SEAs was
0.11. Moreover, the standard deviation of hookworm infection
rates across SEAs in 1910 was 0.23; so a one-standard-deviation
increase in lagged hookworm infection is associated with a post-
RSC increase in schooling enrollment of one quarter of a standard
deviation.

These results are presented in Table II. Estimates of the
variable of interest, ($H_j^{\mathrm{pre}} \times \mathrm{Post}_t$), are displayed for various
outcomes and specifications. Panel A presents the main results.

10. In previous versions of this study, I also compared hookworm infection
with the prevalence of other diseases conditions. I find a relationship between
hookworm and malaria across the counties of the South. On the other hand, I find
no robust relationship between hookworm and child mortality, pellagra morbidity,
or typhoid deaths. These latter two variables were only available for the counties
of one state each, and thus I cannot use them in the subsequent regression
analysis. Malaria mortality rates are used below in the sensitivity analysis.

TABLE II
HOOKWORM AND HUMAN CAPITAL: BASIC RESULTS

| Dependent variables: | | (1) School enrollment | (2) Full-time school attendance | (3) Literacy |
|---|---|---|---|---|
| *Panel A: Basic results* | | | | |
| *Census years* | *Estimating equation* | | | |
| (A) 1910–1920 | (1) | 0.0883*** (0.0225) | 0.1591*** (0.0252) | 0.0587*** (0.0186) |
| (B) 1900–1950 | (1) | 0.0608*** (0.0261) | 0.1247*** (0.0286) | |
| (C) 1900–1950 | (2) | 0.0954*** (0.0233) | 0.1471*** (0.0287) | |
| *Panel B: Effects within and between states* | | | | |
| Change to specification | | | | |
| (D) Include state × Post dummies | | 0.1313*** (0.0245) | 0.2144*** (0.0290) | 0.0417** (0.0207) |
| (E) Allow for state-specific mean reversion | | 0.1148*** (0.0265) | 0.1813*** (0.0312) | 0.0408** (0.0206) |
| (F) Use infection from state of birth instead of SEA | | 0.0489 (0.0504) | 0.2057*** (0.0765) | 0.0907** (0.0451) |
| Census years: | | 1900–1950 | 1900–1950 | 1910–1920 |
| Estimating equation: | | (2) | (2) | (1) |

This table reports estimates of the interaction of pretreatment hookworm and a post-RSC dummy in the indicated equations. The dependent variables are the binary variables denoted in the column headings. Robust standard errors in parentheses (clustering on area times $Post_t$). Single asterisk denotes statistical significance at the 90% level of confidence; double 95%; triple, 99%. All regressions include fixed effects for area and time: controls for age, female, female × age, black, and black × age; and the interactions of the demographic controls with $Post_t$. The average school enrollment in 1910 (× $Post_t$) is used to control for mean reversion in rows (E) and (F). The base sample consists of all native-born white and black children in the IPUMS between the ages of 8 and 16 for 1900, 1910, 1920, 1940, and 1950. For rows (A) through (E), the sample is drawn from the RSC-surveyed county groups (State Economic Areas or SEAs), all of which were in the South. For row (F), the sample consists of individuals in the 50 states and territories for which Kofoid and Tucker [1921] report hookworm infection rates, and all the area-level variables (hookworm, fixed effects, trend, mean-reversion control) are specified at the state-of-birth level. Because literacy is not available in the later Censuses, no estimates are available for literacy in rows (B) and (C).

Row (A) contains the estimates using the 1910 and 1920 censuses, which bracket the RSC intervention, while row (B) contains similar estimates using the Census microdata from 1900–1950. In addition to the results on school enrollment mentioned in the previous paragraph, I estimate positive effects of hookworm eradication on full-time school attendance and literacy as well. (The literacy variable is not available in later Censuses, so Column 3 is
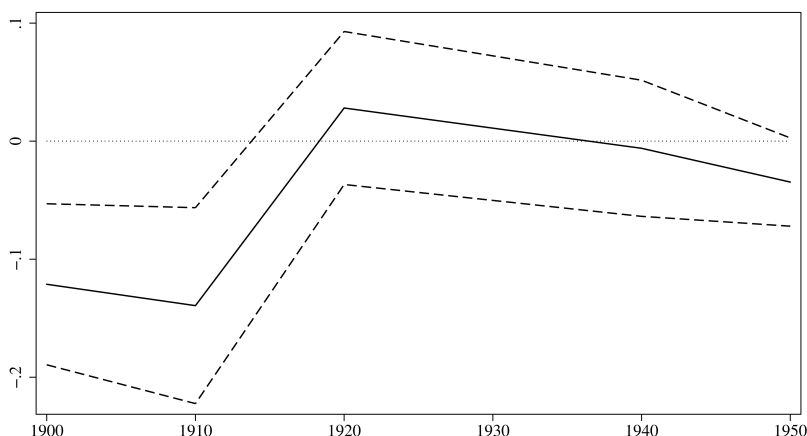
FIGURE II

Hookworm Eradication and School Attendance, 1900–1950

The $y$ axis plots the year-specific coefficients on the circa-1913 hookworm-infection rate (solid line), plus the RSC-confidence intervals (dashed lines). The $x$ axis is the Census year. The sample consists of all native-born white and black children in the IPUMS between the ages of 8 and 16 in the RSC-surveyed geographic units for 1900, 1910, 1920, 1940, and 1950. For each year, the coefficients are estimated in a regression of a school-attendance dummy on preintervention hookworm infection and demographic controls. Confidence intervals are constructed using standard errors that are clustered on SEA.

blank in Rows B and C; literacy results in Panel B use the 1910–20 Censuses.)

The surge in school attendance in high-hookworm counties coincided with the campaign for hookworm eradication. This can be seen in Figure II.[11] As shown in the graph, areas with more hookworm infection had lower levels of school attendance prior to the RSC, but these groups converge markedly thereafter. I further test this hypothesis adding SEA-specific trends to (1). The equation that results is

$$(2) \qquad Y_{ijt} = \beta(H_j^{\text{pre}} \times \text{Post}_t) + \tilde{\delta}_j \times t + \delta_t + \delta_j + X_{ijt}\Gamma + \varepsilon_{ijt}$$

Trend differences across areas will load onto the $\tilde{\delta}_j$, while differences that coincide with the anti-hookworm campaign will load onto $\beta$. Estimates of (2) in row (C) of Table II show little change in the estimated $\beta$.

11. To construct this figure, I run a regression of school attendance on SEA-level hookworm, separately by Census year from 1900 to 1950. Micro level controls for age, female, female×age, black, and black×age are also included. The year-specific estimates on $H_j^{\text{pre}}$ are plotted against year.

The specification in row (D) contains controls for state-level shocks and policy changes, most notably the compulsory-schooling and child-labor laws that were imposed in the first half of the twentieth century. Since these shifts were at the level of state × year, this specification implements a simple fix to purge the estimates of this effect: including (state × year) fixed effects. Throwing out all of the cross-state variation yields estimated effects that are essentially unchanged.

Another concern is mean reversion across areas: if some counties had high hookworm infection and low schooling because of a temporary shock, we might expect rises in school attendance in the following period even if hookworm had not affected the schooling decision. In Panel F of Table II. I add the interaction of $Post_t$ with 1910 average school attendance by SEA the specification. Differential incidence of state policies (by average school-attendance rates) are also absorbed by interacting state × year dummies with average attendance rates. There is evidence of mean reversion in schooling, but estimates of $(H_j^{pre} \times Post_t)$ are similar to above.

In row (F), I re-estimate (2) using only state-of-birth-level variation in the anti-hookworm campaign. Because the RSC did not attempt a systematic survey of hookworm across the whole country, I use hookworm-infection rates from Kofoid and Tucker [1921], who surveyed hookworm among army recruits. Because the full set of states is a much more heterogeneous sample, I also control for mean reversion as above. Restricting the analysis to the state level excludes much of the useful variation: the standard errors on the estimate of $(H_j^{pre} \times Post_t)$ are approximately twice those found above. There are two reasons for this: (i) there are fewer geographic units; and (ii) the dispersion of infection rates across states is smaller than that across county groups.

On the other hand, point estimates of the effect of hookworm eradication are approximately the same magnitude as those in the county-level results. The result for enrollment is smaller than the estimates above, and we can reject neither zero nor the estimates from the SEA-level variation. On the other hand, attendance and literacy do show statistically significant responses to hookworm, with magnitudes that are larger than previous estimates. The results indicate that, at the state-of-birth level, the effect of hookworm eradication worked

principally through the intensive margin of human-capital formation (literacy and full-time school attendance).[12]

### IV.B. Additional Specifications

The finding that highly infected counties experienced surges in school attendance is not sensitive to controlling for a variety of alternative hypotheses. I contrast these hypotheses with the effect of hookworm and the RSC by starting with (1) and (2) and adding plausible proxies for the supposed confounds. The control variables enter into the specification interacted with $Post_t$. These results are found in Panel B of Table III. In every case, the added control variables are jointly significant at conventional confidence levels. The new controls include variables for health and health policy, educational resources, race and race relations, urbanization and land use, and parental background. (See the Appendix for a complete list of controls and their sources.)

The estimated relationship between hookworm and human capital was not simply concentrated on one particular demographic group, although there are noteworthy differences. These results are seen in Panel C of Table III. For preteens and adolescents, the estimates for enrollment and attendance are close in magnitude, which suggests a balancing of two offsetting effects: younger children were more likely to be infected, but adolescents were closer to the margin of not going to school.

There were also important differences between how blacks and whites responded to the anti-hookworm campaign. Whites appeared to have positive responses to hookworm eradication by all three measures of human capital, but, for blacks, the estimated effects of hookworm were uniformly larger. There are several candidate explanations for this result. One is that the general health of blacks was more sensitive to a given level of (own) hookworm infection. However, this explanation is inconsistent with existing medical evidence [Vance 1932]. The other possibility is that whites, because of higher average incomes and

---

12. This suggests that, if we examine these cohorts as adults, we will see increases in human capital, but the estimates (of years of schooling especially) may well be statistically insignificant. These results provide a natural benchmark for the cohort-based analysis in Section VI below, since the retrospective-cohort analysis employs precisely the state of-birth variation in hookworm.

TABLE III

HOOKWORM AND HUMAN CAPITAL: SENSITIVITY TESTS AND RESULTS FOR SUBGROUPS

| Dependent variables: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | School enrollment | | Full-time school attendance | | Literate |
| Censuses: | 1900–50 | 1910–20 | 1900–50 | 1910–20 | 1910–20 |
| *Specification or subsample:* | | | | | |
| | *Panel A: Baseline results* | | | | |
| Baseline | 0.0954*** | 0.0883*** | 0.1471*** | 0.1591*** | 0.0587*** |
| | (0.0233) | (0.0225) | (0.0287) | (0.0252) | (0.0186) |
| | *Panel B: Specifications with additional controls* | | | | |
| Health and health policy | 0.1200*** | 0.1187*** | 0.1628*** | 0.1646*** | 0.0724*** |
| | (0.0291) | (0.0262) | (0.0355) | (0.0294) | (0.0233) |
| Education and race | 0.1235*** | 0.0793*** | 0.1851*** | 0.1581*** | 0.0556*** |
| | (0.0208) | (0.0208) | (0.0247) | (0.0250) | (0.0171) |
| Full controls | 0.1014*** | 0.0850*** | 0.1408*** | 0.1026*** | 0.0513** |
| | (0.0349) | (0.0224) | (0.0421) | (0.0325) | (0.0213) |
| | *Panel C: Demographic subgroups* | | | | |
| Preteens | 0.0932*** | 0.0890*** | 0.1416*** | 0.1549*** | 0.0912*** |
| | (0.0255) | (0.0242) | (0.0302) | (0.0266) | (0.0253) |
| Adolescents | 0.0986*** | 0.0877*** | 0.1573*** | 0.1682*** | 0.0323* |
| | (0.0280) | (0.0282) | (0.0336) | (0.0295) | (0.0165) |
| Blacks | 0.2299*** | 0.1838*** | 0.2601*** | 0.2205*** | 0.1078*** |
| | (0.0399) | (0.0337) | (0.0399) | (0.0320) | (0.0374) |
| Whites | 0.0378 | 0.0270 | 0.1103*** | 0.1169*** | 0.0264* |
| | (0.0237) | (0.0267) | (0.0294) | (0.0294) | (0.0139) |

This table reports estimates of the interaction of pretreatment hookworm and a post-RSC dummy in (1) (for the 1910–20 data) and (2) (for the 1900–50 data), for the indicated subsamples. The dependent variables are the binary indicators denoted in the column headings. Robust standard errors in parentheses (clustering on SEA times post). Single asterisk denotes statistical significance at the 90% level of confidence; double, 95%; triple, 99%. Sample consists of native-born black and white children in the IPUMS between the ages of 8 and 16 in the RSC-surveyed geographic units for the indicated years. The aggregate control variables enter into the specification interacted with $Post_t$. Control variables are described in the Appendix. Number of clusters = 230. All regressions include fixed effects for area and time; controls for age, female, female × age, black, and black × age; and the interactions of the demographic controls with $Post_t$. Reporting of additional coefficient estimates is suppressed.

therefore better sanitary conditions, had lower rates of infection. Unfortunately, there is no direct published evidence on this hypothesis. A third explanation is that whites, who were more likely than blacks to go to school and be literate, simply had less scope to increase much along these measures of human-capital invest-

ment.[13] The long-term consequence of these racial differences is less clear because the return to schooling was lower for blacks than whites during this period. I revisit this issue in Section VI.

### IV.C. Interpretation

The estimates presented above imply plausible numbers for the effect of hookworm infection on school attendance. We can compare the reduced-form effect of $(H_j^{\text{pre}} \times \text{Post}_t)$ (about 0.09) to the estimated decline in infection as a function of the same variable (0.44). The latter number comes from the resurveys discussed in Section II.D, and is shown in Figure I. Some of this relationship may be due to "Galton's fallacy" since it is a comparison of $\Delta H_t$ and $H_{t-1}$. This resulting upward bias in the first-stage relationship will cause a downward bias in the indirect-least-squares (ILS) estimates below.[14] On the other hand, the intervention may have reduced the rate of severe infections more than the overall rate, which would likely cause an upward bias in this estimator. Furthermore, in spite of the extensive set of controls employed above, it is impossible to rule out that, for example, the RSC intervention differentially improved other sanitation-related diseases for hookworm-infested areas. Again, this presumably results in an upward bias on the ILS estimator.

Dividing the first number by the second gives us the indirect least squares estimate of infection on enrollment: 0.20. This indicates that a child infected with hookworm is 20 percent (i.e., percentage points) less likely to be enrolled. Similarly, ILS estimates imply a 0.13 lower probability of being literate and 0.33 reduction in the probability of attending school full time.[15]

These estimates suggest that hookworm played a major role in the South's lagging behind the rest of the country. In comput-

13. Using logit and probit estimators rather a linear probability model, I find that the hookworm effect is sometimes larger for whites, sometimes not, depending on the specification. This bolsters the hypothesis that the two groups are experiencing similar increases in some latent measure (human-capital investment), but that binary nature of the census variables obscures this to some degree.

14. The mean-reversion bias on the reduced-form coefficients was shown to be negligible in Section IV. A, so there should be no mean-reversion bias in the ILS numerator.

15. These numbers suggest a larger effect than those obtained by Miguel and Kremer [2004], who report an IV estimate of $-0.203$ for the effect of intestinal-parasite infection on school participation. Their variable is based on spot checks of school attendance following the intervention, and therefore is most comparable to the full-time school attendance variable in the present study. However, the estimates are not directly comparable because Miguel and Kremer use a combined infection rate that includes hookworm, roundworm, schistosomiasis, and whipworm.

TABLE IV
CONTEMPORANEOUS EFFECT ON ADULT OUTCOMES

| Samples: | (1) Whole | (2) Male | (3) Female | (4) White | (5) Black |
|---|---|---|---|---|---|
| | Parameter estimates | | | | |
| *Dependent variables:* | | | | | |
| Literacy | 0.0062 | −0.0107 | 0.0203 | 0.0107 | −0.0014 |
| | (0.0095) | (0.0108) | (0.0127) | (0.0112) | (0.0229) |
| Labor-force | −0.0069 | −0.0069 | −0.0056 | −0.0212 | 0.0036 |
| participation | (0.0134) | (0.0065) | (0.0284) | (0.0124) | (0.0249) |
| Occupational income | 0.0526 | −0.0186 | 0.0581 | 0.0855 | 0.0224 |
| score | (0.2836) | (0.4912) | (0.4163) | (0.3903) | (0.3861) |
| Lives in an urban area | 0.0157 | 0.0030 | 0.0280 | 0.0199 | 0.0132 |
| | (0.0172) | (0.0190) | (0.0177) | (0.0226) | (0.0245) |

Each cell reports the coefficient estimate on Hookworm × Post for the indicated sample and dependent variable. Robust standard errors in parentheses (clustering on SEA times post, number of clusters = 230). None of the reported coefficients is statistically significant at conventional confidence intervals. The sample consists of all native-born white and black adults in the 1910–1920 IPUMS between the ages of 25 and 55 (inclusive) in the RSC-surveyed geographic units. Reporting of additional coefficient estimates is suppressed. Specifications also include dummy variables for SEA, age, black, female, and year, as well as interactions of the demographic variables with $Post_t$.

ing the depressing effect of hookworm on the region's human-capital accumulation, I multiply the ILS estimates from above with an estimate of the area's hookworm burden. I assume a 40 percent regional hookworm-infection rate, as reported by the RSC. The resulting numbers account for around half of the human-capital gap.

## V. CONTEMPORANEOUS EFFECTS ON ADULTS

Next, I examine how adult outcomes in the same time periods respond to the anti-hookworm campaign. This serves as a falsification exercise because adults were less likely to be directly affected by eradication. As an empirical matter, adults had much lower infection rates.[16]

Results for adults are displayed in Table IV, which contains estimates of (1). For several outcomes, I cannot reject the null hypothesis that there was no differential change across counties

16. Smillie and Augustine [1925] show that hookworm infection among adults was very low in the Southern United States. They also note the contrast with the experience of other countries, where hookworm infects across a broader range of ages.

with different hookworm-infection rates. Neither adult literacy nor labor-force participation was significantly affected by the treatment campaign. I obtain null results for the effect of hookworm on the occupational income score, an IPUMS variable that proxies income by occupation. I also do not find evidence that adults were more likely to live in urban areas. Finally, in results not shown, $(H_j^{\text{pre}} \times \text{Post}_t)$ does not predict whether adults residing in the area were in white-collar jobs or born out of state, nor do I find significant effects when I perform the above analysis separately by broad age groupings.

## VI. Long-Term Follow-Up of Cohorts Exposed as Children

In this section, I follow up on the subsequent outcomes of the cohorts that, as children, were exposed to the hookworm-eradication campaign. This analysis therefore represents a different approach to the question: instead of looking at the behavior of fixed age groups at different points in time, I analyze various year- and state-of-birth cohorts retrospectively. The comparisons are both across areas, based on different preexisting infection rates, and across cohorts, with older cohorts serving as a comparison group because they were not exposed to the RSC during childhood.

The geographic units employed in this analysis are place of birth rather than current residence. Matching individuals with hookworm-infection rates of the area where they end up as adults would be difficult to interpret because of migration. Instead, I use the information on hookworm prevalence in an individual's state of birth to conduct the analysis. A problem with using states instead of counties is that there are fewer of them. As seen above, this reduces precision.

The effects of hookworm infection among children appeared to extend into adulthood for the affected cohorts. This section contains several results supporting this conclusion.

### VI.A. Results for Earnings, Schooling, and Literacy

I consider a simple parameterization of the cross-cohort comparison: the number of *childhood* years potentially exposed to the anti-hookworm campaign, times the pre-eradication hookworm intensity in the state of birth. Exposure to the RSC, $\text{Exp}_{ik}$, will be zero for older cohorts, rise linearly for those born in the nineteen

years prior to 1910, and stop at nineteen for younger cohorts.[17] Nineteen is chosen because most individuals in this period would have completed their schooling by that age, and hookworm infection was negligible at older ages. Thus, the regression model is

$$(3) \qquad Y_{ijk} = \beta(H_j^{\text{pre}} \times \text{Exp}_{ik}) + \delta_j + \delta_k + X_{ijk}\Gamma + v_{ijk}$$

for state of birth $j$ and cohort $k$. The demographic controls consist of indicator variables for each age $\times$ black $\times$ female cell, plus interactions of state-of-birth dummies with black, female, and black $\times$ female.

Children with more exposure to the campaign, by being born later and in a state with greater pre-eradication hookworm, were more likely to be literate and earn higher incomes as adults. Results are mixed for years of schooling, but this is within the range of normal statistical variation. Table V contains these results. Panel A presents the estimates of (3).

The estimates do not appear to be an an artifact of mean reversion. If the oldest cohorts had high hookworm infection and low productivity because of some mean-reverting shock, we might expect income gains for the subsequent cohorts even in the absence of a direct effect of hookworm on productivity. I use data on labor earnings by state in 1899 from Lebergott [1964]. I interact the natural logarithm of this measure with age and include the interaction in the even-numbered columns of Table V. This analysis yields mixed evidence of mean reversion in the data, but the inclusion of these controls does not substantially affect the coefficient on $H_j^{\text{pre}} \times \text{Exp}_{ik}$.[18]

I argue that the earnings results are not contaminated by hookworm-induced changes in the probability of self-employment. The major difficulty in using the earnings data from the 1940 Census is that it was incomplete: labor income from self-employment was excluded. To gauge the impact of this problem, I estimate regression equations identical to those used for Panel A of Table V, but with three new dependent variables: binary indicators for (i) self employment, (ii) missing data for log earnings, and (iii) non-wage/salary income being greater than fifty dollars. In doing so, I find no robust and statistically significant

17. Specifically, the formula is $\text{Exp}_{ik} = \max(\min(19, 49 - \text{age}_i), 0)$ where age is measured in 1940.
18. Similar results are obtained by interacting the 1899 wage measure with the exposure variable instead of age and/or including the square of the average wage ($\times$ age) as well.

TABLE V

LONG-TERM FOLLOWUP BASED ON INTENSITY OF EXPOSURE TO THE TREATMENT CAMPAIGN

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Controls for mean-reversion: | No | Yes | No | Yes | No | Yes |
| Dependent variables: | Log earnings, 1939 | | Years of schooling, 1940 | | Literacy status, 1920 | |
| | | | | | | |
| *Panel A: Main results* | | | | | | |
| *Independent variables* | | | | | | |
| Hookworm infection Rate × years of exposure | 0.0286*** | 0.0234** | −0.0243 | 0.0037 | 0.0158*** | 0.0115*** |
| | (0.0066) | (0.0093) | (0.0328) | (0.0357) | (0.0019) | (0.0020) |
| | | | | | | |
| *Panel B: Changing returns to schooling* | | | | | | |
| *Independent variables* | | | | | | |
| Hookworm infection Rate × Years of exposure | 0.0254*** | 0.0219*** | n.a. | | n.a. | |
| | (0.0044) | (0.0063) | | | | |
| Infection × Years of exposure × Years of schooling | 0.0023*** | 0.0022** | | | | |
| | (0.0009) | (0.0009) | | | | |
| | | | | | | |
| *Panel C: Estimates of hookworm × exposure for demographic subgroups* | | | | | | |
| *Subsamples* | | | | | | |
| Males | 0.0265*** | 0.0253*** | −0.0690** | −0.0376 | 0.0108*** | 0.0083*** |
| | (0.0056) | (0.0080) | (0.0326) | (0.0347) | (0.0018) | (0.0019) |
| Females | 0.0322*** | 0.0157 | 0.0200 | 0.0444 | 0.0209*** | 0.0148*** |
| | (0.0115) | (0.0165) | (0.0338) | (0.0385) | (0.0027) | (0.0030) |
| Whites | 0.0293*** | 0.0232** | −0.0110 | 0.0164 | 0.0131*** | 0.0086*** |
| | (0.0071) | (0.0103) | (0.0345) | (0.0378) | (0.0022) | (0.0020) |
| Blacks | 0.0220*** | 0.0253** | 0.1013*** | 0.0133 | 0.0314*** | 0.0262*** |
| | (0.0072) | (0.0103) | (0.0387) | (0.0461) | (0.0065) | (0.0063) |

Each panel/column reports reports a separate regression for the indicated samples and dependent variables. State-average data are matched to individuals based on their state of birth. The measure of hookworm is from Kofoid and Tucker [1921]. Unskilled-wage data from 1899, reported by Lebergott [1964], are used to control for mean reversion. The full sample consists of native-born blacks and whites in the age range [25, 60] and in the 1940 IPUMS database (except the literacy regressions, which include ages [16, 60] from the 1920 IPUMS data). Robust standard errors in parentheses (clustering on state of birth). Single asterisk denotes statistical significance at the 90% level of confidence: double, 95% triple 99%. The demographic controls consist of indicator variables for each age × black × female cell, plus interactions of statu-of-birth dummies with black, female and black × female. Reporting of additional coefficient estimates is suppressed.

relationship between the hookworm measure and any of these three measures.[19] Additionally, I find evidence of hookworm-related increase in the total time worked (either for wage/salary or not), although, once mean reversion controls are included, the increased labor supply does not account for a large fraction of the earnings effect.

I also consider the role played by the quantity of and returns to schooling in the wage results. Controlling directly for education does not significantly change the estimated effect of hookworm treatment. Additionally, I can easily reject, for conventional returns to schooling, the hypothesis that the wage effect is due entirely to a rise in education.[20] However, the fact that I estimate increases in literacy without concomitant rises in the *quantity* of schooling suggests an alternative hypothesis: changes in quality. In particular, it may be that students spend the same number of years in school, but the time is better spent. For example, there might be less absenteeism, or students might be better equipped to absorb the material while in school. As was shown above, students were less likely to work while in school and were more more likely to be literate, following hookworm eradication. This suggests that the return to schooling was raised by the hookworm intervention.

There is indeed evidence that the return to schooling rose with the intervention. This can be seen in Panel B of Table V. The crucial interaction is the triple: between years of schooling and the treatment-intensity variable ($H_j \times \text{Exp}_{ik}$).[21] This new term is estimated to be positive and statistically significant in the labor-earnings regression.

This hookworm-related change in the return to schooling can potentially explain a large fraction of the increase in earnings

19. Using controls for mean reversion, the coefficient (standard error) on $H_j \times \text{Exp}_{ik}$ is estimated to be $-0.003$ (0.003) for self-employment, 0.003 (0.004) for missing log-earnings data, and $-0.001$ (0.004) for non-wage/salary income exceeding fifty dollars.

20. Similarly, I do not find evidence that the mechanism is migration out of the South, migration into an urban area, shifting to white-collar occupations, or movement out of agriculture. I consider these potential channels by conditioning on the variables in the regressions above, and find that their inclusion makes little difference for the estimate of $H_j \times \text{Exp}_{ik}$. Such variables are themselves endogenous, and therefore these results should be considered a decomposition of the hookworm effect, taking the regression estimates of the added variables as correct.

21. The additional second-order interactions are absorbed with a series of dummies for birth state × education and birth year × education. The first-order effects of education, state of birth, and year of birth are also absorbed with indicator variables.

described above. This regression, by comparing individuals with different terminal levels of attainment, estimates the average marginal effect of schooling in the sample (and how it changes following hookworm eradication). If the intervention had similar effects on the return to inframarginal schooling, we can compute the overall contribution through this channel. Multiplying the triple interaction (0.0022) by the average years of schooling in the South (7.72) yields 0.0170, almost 80 percent of the coefficient on $(H_j \times \mathrm{Exp}_{ik})$ in the first row of Panel B.[22] Moreover, I cannot reject the hypothesis that all of the earnings effects worked through the rising return to schooling.

Several differences emerge among demographic groups. Results estimated from subsamples of males, females, whites, and blacks are contained in Table V, Panel C. For no subgroup is there a robustly significant relationship between years of schooling and $(H_j \times \mathrm{Exp}_{ik})$. Estimates for literacy, on the other hand, are positive and significantly different from zero for all demographic groups. Literacy responses are larger for females than for males, as well as for blacks than for whites, possibly because females and blacks had lower preexisting literacy rates. The estimate of $(H_j \times \mathrm{Exp}_{ik})$ in the earnings equation yields a positive and significant number for males, while the result for females is not sensitive to the inclusion of the mean-reversion control. Whites and blacks show similar earnings responses, on the other hand. This may be because blacks, while gaining more on measured human capital, faced lower returns to skill in the labor market.

### VI.B. Cohort-Specific Relationship Between Income and Hookworm

In this subsection, I show that the shift in the relationship between income and pre-eradication hookworm coincides with childhood exposure to the eradication efforts. This can be seen graphically, and I also provide statistical tests comparing exposure to eradication with trends and or autoregressive processes.

I use two income proxies that are available for a large number of Censuses. The occupational income score and Duncan socioeconomic index are both average indicators by disaggregated occupational categories that were calibrated using data from the 1950 Census. The former variable is the average by occupation of all reported labor earn-

---

22. Because the education variable was de-meaned before interaction, the second-order term is evaluated at the mean of education.

ings. The measure due to Duncan [1961] is instead a weighted average of earnings and education among males within each occupation. Both variables can therefore measure shifts in income that take place between occupations. The Duncan measure has the added benefit of picking up between-occupation shifts in skill requirements for jobs. Occupation has been measured by the Census for more than a century, and so these income proxies are available for a substantial stretch of cohorts.

Using these proxies, I construct a panel of average income by cohort. The units of observation of the panel are year of birth $\times$ state of birth, and I use microsamples from ten Censuses (1880 and 1900–1990). This results in an unbalanced panel spanning the year-of-birth cohorts from 1825 to 1965 for forty six states of birth. (See the Appendix for the details of the data construction.)

For each year of birth, OLS regression coefficients are estimated on the resulting cross section of states of birth. Consider a simple regression model of an average outcome, $Y_{jk}$, for a cohort with state of birth $j$ and year of birth $k$:

$$(4) \qquad Y_{jk} = \beta_k H_j^{\text{pre}} + \delta_k + X_j' \Gamma_k + \nu_{jk}$$

in which $\beta_k$ is year-of-birth-specific coefficient on hookworm, $X_j$ is a vector of other state-of-birth controls, and $\delta_k$ and $\Gamma_k$ are year-of-birth-specific intercept and slope coefficients. (Note that there is no $i$ subscript because I am working with average outcomes by cohort.) I estimate this equation using OLS for each year of birth $k$. This specification allows us to examine how the relationship between income and pre-eradication hookworm ($\hat{\beta}_k$) differs across cohorts.

I start with a simple graphical analysis using this flexible specification for cross-cohort comparison. Figure III displays a plot of the estimated $\beta_k$. Each cohort's point estimate is marked with a dot. The top row of graphs contains estimates from the basic specification, in which the average income is regressed on hookworm infection, Lebergott's measure of 1899 wage levels, and dummies for Census region. The bottom row displays estimates from the "full controls" specification, which, in addition to the basic variables, contains a number of control variables.[23]

---

23. These consist of the following state-of-birth-level variables: 1910 fraction black, fraction literate (among adults), fraction of population living in urban areas, 1890 child-mortality rate, fraction of deaths in 1890 caused by scarlet fever, measles, whooping cough, diphtheria/croup, typhoid fever, malaria, diarrheal diseases, and pneumonia, 1910 fertility rates, 1930 unemployment rate, doctors per capita in 1898, state public health spending per capita in 1898, WWI recruits
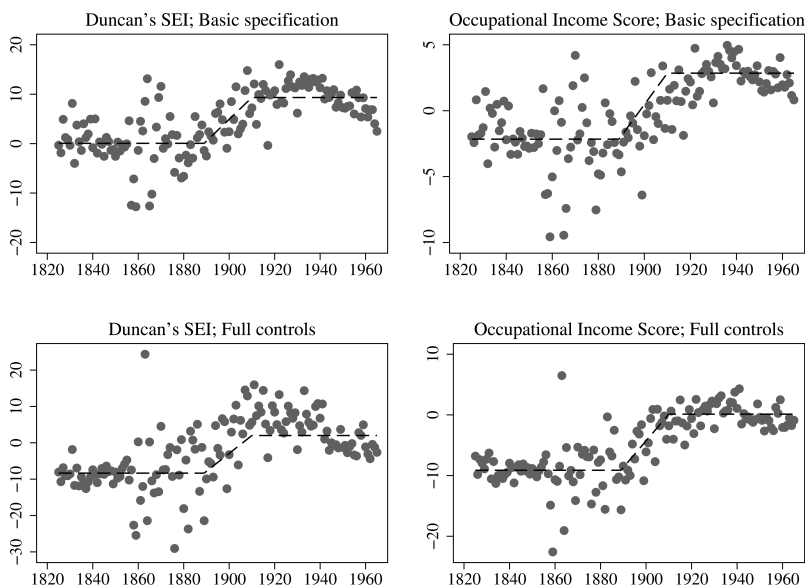
FIGURE III

Cohort-Specific Relationship Between Income and Pre-Eradication Hookworm

These graphics summarize regressions of income proxies on pre-eradication hookworm-infection rates by state of bith. The *y* axis for each graphic plots the estimated cohort-specific coefficients on the state-level hookworm measure. The *x* axis is the cohort's year of birth. Each year-of-birth cohort's point estimate is marked with a dot. The dashed lines measure the number of years of potential childhood exposure to the Rockefeller Sanitary Commission's activities. For the undertaking regressions, the dependent variables are constructed from the indicated income proxies (the Duncan Socioeconomic Indicator and the Occupational Income Score). For each year-of-birth cohort, OLS regression coefficients are estimated on the cross section of incomes by state of birth. In the basic specification, this state-of-birth average income is regressed onto hookworm infection, Lebergott's measure of 1809 wage levels, and regional dummies. The "full controls" specification contains in addition the various controls variables described in the Appendix. The regressions are estimated using weight equal to the square root of the cell size in the underlying microdata.

Because hookworm was principally a childhood disease, cohorts that were already adults in 1910 were too old to have benefited from the reduction in hookworm. On the other hand, later cohorts experienced reduced hookworm infection during their childhood. This benefit increased with younger cohorts who

---

found "defective" at draft physical, and 1902–1932 logarithmic changes in average monthly teacher salaries, school term length, school expenditures per capita, and pupil/teacher ratios. See the Appendix for details on these data.

were exposed to the RSC's efforts for a greater fraction of their childhood. The dashed lines therefore measure the number of years of potential childhood exposure (defined above) to the Rockefeller Sanitary Commission's activities.[24] Pre-eradication hookworm generally predicts lower income in earlier cohorts, while this is not the case for those born late enough to have potentially been exposed to the RSC during childhood.

Formal statistical tests indicate that the shift in the income/pre-eradication–hookworm relationship coincided with exposure to hookworm eradication, rather than with some trend or autoregressive process. I treat the $\hat{\beta}_k$ as a time series and estimate the following regression equation:

$$(5) \qquad \hat{\beta}_k = \alpha \mathrm{Exp}_k + \sum_{i=1}^{n} \gamma_n k^n + \Phi(L)\hat{\beta}_k = \eta_{ljk}$$

in which $\mathrm{Exp}_k$ is exposure to hookworm eradication (defined similarly above), the $k^n$ terms are $n$th-order trends, and $\Phi(L)$ is a distributed lag operator. To account for the changing precision with which the generated observations are estimated, observations are weighted by the inverse of the standard error for $\hat{\beta}_k$. Table VI reports estimates of (5) under a variety of order assumptions about trends and autoregression. The dependent variables are the $\hat{\beta}_k$, from the specification using the broad sets of controls. For the analysis using the occupational income score, the exposure term is similar across specifications, and there is no statistically significant evidence of trends or autoregression in these $\hat{\beta}_k$. The estimates using the Duncan SEI exhibit both trend and autoregression, but the exposure coefficient is stable once both are accounted for.[25]

---

24. The exposure variable is measured in different units than the cohort-specific regression coefficients, so, to facilitate visual comparisons in Figure III, the line is rescaled in the $y$ dimension so that the pre-1880 and post-1930 levels match those of the $\beta_k$. On the other hand, the calendar years for childhood exposure are fixed by the start of the RSC and the observed life-cycle pattern of infection, and therefore the exposure line is not rescaled in the $x$ dimension.

25. I obtain similar results from a variety of alternative specifications and methodologies for constructing the income data. These include constructing the cohort income data from narrower age ranges, excluding the 1990 Census, or including the 2000 Census. The results are not sensitive to using an unweighted specification. Moreover, restricting the analysis to one hundred and then forty years of birth cohorts yields similar estimates on exposure, albeit with larger standard errors. On the other hand, I have experimented with higher-order polynomial trends and found no estimates of exposure that are statistically significant for n ≤ 5.

TABLE VI

EXPOSURE TO RSC VERSUS ALTERNATIVE TIME-SERIES RELATIONSHIPS

| Income proxy | Specification | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|---|
| | | Time-series estimates of the exposure coefficient | | | | |
| Occupational income score | Basic | 0.3113*** | 0.2915*** | 0.2612*** | 0.2497*** | 0.1912*** |
| | | (0.0214) | (0.0542) | (0.0384) | (0.0612) | (0.0622) |
| Occupational income score | Full controls | 0.2623*** | 0.3732*** | 0.2346*** | 0.3393*** | 0.2743*** |
| | | (0.0339) | (0.0858) | (0.0438) | (0.0960) | (0.1007) |
| Duncan's socioeconomic indicator | Basic | 0.5352*** | 0.7566*** | 0.3928*** | 0.5983*** | 0.4858*** |
| | | (0.0418) | (0.1069) | (0.0520) | (0.1124) | (0.1282) |
| Duncan's socioeconomic indicator | Full controls | 0.5007*** | 0.8820*** | 0.3544*** | 0.6616*** | 0.7081*** |
| | | (0.0661) | (0.1707) | (0.0735) | (0.1791) | (0.1969) |
| *Additional controls* | | | | | | |
| Order of Polynomial Trend: | | 0 | 1 | 0 | 1 | 2 |
| Order of Autoregressive Process: | | 0 | 0 | 1 | 1 | 2 |

This table reports estimates of childhood exposure to the RSC in (5). The dependent variables are the cohort-specific regression estimates of income proxies on hookworm that are shown in Figure III. Robust standard errors are given in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence; double 95%, triple, 99%. Observations are weighted by the inverse of the standard error for $\beta_k$. In the basic specification, the income proxy was regressed on to hookworm infection, Lebergott's measure of 1899 wage levels, and dummies for the the four census regions. The "full controls" specification contains, in addition, the various control variables listed in the Appendix.

## VI.C. Interpretation

In this section, I characterize the magnitude of the effect of the hookworm reduction in more easily interpretable units, and contrast the estimates with cross-area differences in income per capita. I focus on the contrast between the cohort with zero childhood exposure to the RSC and the cohort with full exposure. For example, comparing these cohorts in two areas that were one standard deviation (within the RSC-targeted area) apart in hookworm, we would expect wages to have increased 11 percent more in the area with the higher pre-period infection rate.

Using indirect least squares (ILS), I estimate the approximate effect of childhood hookworm infection on adult wages to be around 43 percent. Again, I compare the fully exposed and nonexposed cohorts to construct the estimate. The increase in wages as a function of the Kofoid measure of $H_j^{\text{pre}}$ is 0.32, when comparing zero to full RSC exposure. Since hookworm was largely eradicated in the time span considered, I regress the pre-RSC hookworm (reported by Jacocks at the state level for certain states) on the Kofoid measure and estimate that the decrease in infection rates as a function of $H_j^{\text{pre}}$ to be 0.748.[26] This yields an ILS estimate of $-0.43$ in natural-log terms.

The ILS estimates using the occupational proxies for income are similar to the wage result. The estimated shift in income related to RSC exposure was estimated in Table VI. I compare these estimated changes with their respective averages for men born in the South between 1875 and 1895, and then construct the ILS coefficient as above. For the occupational income score. I estimate that the proportional change in income related to childhood hookworm infection is $-0.23$. The same estimate for the Duncan SEI is $-0.42$.

These results point to changes in the returns to schooling as well. I compute the drop in returns due to childhood hookworm

26. This is a departure from the methodology in Section IV.C in that I scale the reduced-form coefficient by the pre-existing hookworm infection rate, rather than the change. For the ILS calculation in Section IV, I used follow-up data on infection rates several years after the RSC to gauge the first-stage relationship between pre-RSC hookworm and the decline. In contrast, I consider in this section the effect of the intervention over a span of many more years, by which line hookworm had been mostly eradicated. Because eradication was slightly less than complete, this will induce in a slight downward bias of the ILS estimate. On the other hand, as in the previous ILS calculation, if the intervention decreased the severity of infections more than the overall rate or the control strategy above did not correct for the correlation between hookworm eradication and other improvements in sanitation, there would be an upward bias in this estimator.

exposure to be approximately 0.047, the ILS estimate for the changing returns to a year of education. This represents a substantial drop due to hookworm—around 50 percent of the estimated return to schooling in this period.

The estimated impact is large enough that it bears consideration in a macroeconomic context, although it is not so large that it unreasonably explains everything. The log income gap between the North and the South at 1900 was approximately 0.75. For a 40 percent infection rate in the South and an effect of hookworm on wages of 0.43, we would expect a reduction in Southern incomes of approximately seventeen log percentage points. In other words, some 22 percent of this income gap could be attributed to hookworm infection in the South. On the other hand, if we turn to contemporaneous evidence from developing countries, Miguel and Kremer estimate that the prevalence of intestinal-parasitic infection in the Busia region of Kenya is around 90 percent among school-aged children. Applying our estimates to this area of Kenya suggests a log-income gain of approximately 0.38 from a complete eradication of intestinal worms from the country. This would be enough to raise income per capita to match the level of Zimbabwe, but obviously well short of the almost three natural-log points needed to reach OECD levels. In any case, this calculation is probably optimistic in that the improvement in health in the U.S. South was translated into economic benefits in part because of functioning institutions like schools and labor markets that could effectively channel the new human capital. Whether this would happen in Kenya is less clear.

## VII. Conclusion

This study evaluates the economic consequences of the successful eradication of hookworm in the American South. The advantages of evaluating this intervention are that (i) its timing was relatively short and well defined, (ii) geographical differences in infection permit a treatment/control design, and (iii) sufficient time has passed that we can evaluate its long-term consequences.

I find that areas with higher levels of hookworm infection prior to the RSC experienced greater increases in school attendance and literacy after the intervention. This result is robust to controlling for a variety of alternative hypotheses, including differential trends across areas, changing crop prices, and shifts in certain education policies. No significant results are found for the

sample of adults, who should have benefited less from the intervention owing to their substantially lower (prior) infection rates. Moreover, a long-term follow-up of affected cohorts indicates a substantial income gain as a result of the reduction in hookworm infection. This follow-up also shows a marked increase in the quality rather than the quantity of education.

This study contributes to two important questions in the literature. One is historical: did the reduction in the relative disease burden play a role in the subsequent convergence between the American North and South? Above I show that the hookworm infection rate could account for around half of the literacy gap and about 20 percent of income differences, and so eradication would have closed it by a similar amount. Another question is contemporary: How much does disease contribute to underdevelopment in the tropics?[27] The present study suggests potentially large benefits of public-health interventions in developing countries, where hookworm is still endemic today. Nevertheless, I show using a simple calculation that, while reducing hookworm infection could bring substantial income gains to some countries, the estimated effect is approximately an order of magnitude too small to be useful in explaining the global income distribution.

While this broad decomposition of income per capita into institutions versus geography is interesting, one might argue that social scientists should instead focus on the efficacy of specific interventions. Changing the geography or the colonial history of a country is impossible, and unfortunately the institutions literature has little to say about the complicated mess of intermediate variables that determine productivity. The present study quantifies the benefits of one such intervention and finds them to be substantial.

Nevertheless, it remains an open question whether the long-term gains from hookworm eradication estimated for the American South can be realized for developing countries in the present day. As noted above, there have been other episodes in which externally supported eradication efforts failed because of a lack of local follow-up. Moreover, even if eradication could be achieved in less-developed areas, presumably a whole range of institutional

---

27. This possibility has been advanced recently by Jeffrey Sachs [2001] as part of an agenda highlighting the importance of geographic factors in development. This view has been challenged by Acemoglu, Johnson, and Robinson [2001] who argue in favor of the importance of institutions over geographic determinism.

infrastructure (functioning schools not least among them) needs to be in place to take advantage of the improvement in health. Investigating these interactions between health and institutions is therefore an important avenue of future study.

UNIVERSITY OF CHICAGO, GRADUATE SCHOOL OF BUSINESS

## REFERENCES

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review,* XCI (2001), 1369–1401.
Bleakley, Hoyt, "Malaria and Human Capital: Evidence from the American South," Mimeo, Massachusetts Institute of Technology, May 2002a.
——, "Three Empirical Essays on Investment in Physical and Human Capital," PhD dissertation. Massachusetts Institute of Technology, 2002b.
——, "Disease and Development: Evidence from the American South," *Journal of the European Economic Association,* I (2003), 376–386.
——, and Fabian Lange, "Chronic Disease Burden and the Interaction of Education, Fertility and Growth," Mimeo, Yale University, 2004.
Brinkley, Garland L., "The Economic Impact of Disease in the American South, 1860–1940," PhD dissertation, University of California, Davis, 1994.
Dickson, Rumona, Shally Awasthi, Paula Williamson, Colin Demellweek, and Paul Garner, "Effects of Treatment for Intestinal Helminth Infection on Growth and Cognitive Performance in Children: Systematic Review of Randomised Trials," *British Medical Journal,* CCCXX (2000), 1697–1701.
Duncan, Otis D., "A Socioeconomic Index for All Occupations," *Occupations and Social Status,* A. J. Reiss, ed. (New York, NY; The Free Press, 1961), 109–138.
Ettling, John, *The Germ of Laziness: Rockefeller Philanthropy and Public Health in the New South,* Cambridge, MA: Harvard University Press, 1981.
Farmer, Henry F., Jr., "The Hookworm Eradication Program in the South. 1909–1925," PhD dissertation, University of Georgia, 1970.
Ferrell, John A., Wilson G. Smillie, Platt W. Covington, and Pauline A. Mead. "Health Departments of States and Provinces of the United States and Canada," *Public Health Bulletin,* CLXXXIV (1932), pp. 1–74.
Havens, Leon C., and Ruth Castles. "The Evaluation of the Hookworm Problem of Alabama by Counties," *Journal of Preventive Medicine,* IV (1930), 109–114.
Jacocks, W. P., "Hookworm Infection Rates in Eleven Southern States as Revealed by Resurveys in 1920–1923," *Journal of the American Medical Association,* LXXXII (1924), 1601–1602.
Kofoid, Charles A., and John P. Tucker, "On the Relationship of Infection by Hookworm to the Incidence of Morbidity and Mortality in 22,842 Men of the United States Army," *American Journal of Hygiene,* I (1921), 79–117.
Kremer, Michael, and Edward Miguel, "The Illusion of Sustainability," Working Paper 10324. Cambridge, MA: National Bureau of Economic Research, 2004.
Lebergott, Stanley, *Manpower in Economic Growth: The American Record Since 1800.* (New York, NY: McGraw-Hill, 1964).
Miguel, Edward, and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica,* LXXII (2004), 159–217.
North Atlantic Population Project, NAPP: Complete Count Microdata, Preliminary Version 0.2. Minneapolis, MN: Minnesota Population Center. Computer file, 2004, (http://www.nappdata.org/). [Accessed November 12, 2005.]
Philipson, Tomas J., "External Treatment Effects and Program Implementation Bias," Working Paper T0250. Cambridge, MA: National Bureau of Economic Research, 2000.
Rockefeller Sanitary Commission. Annual Reports, New York, N.Y.: Rockefeller Foundation (1910–1915).
Ruggles, Steven, and Matthew Sobek, Integrated Public Use Microdata Series:

Version 2.0. Minneapolis, MN: Historical Census Projects, University of Minnesota, 1997, (http://www.ipums.umn.edu/).
Sachs, Jeffrey D., "Tropical Underdevelopment," Working Paper 8119, Cambridge, MA: National Bureau of Economic Research, 2001.
Smillie, Wilson G., and Donald L. Augustine. "Intensity of Hookworm Infection in Alabama: Its Relationship to Residence, Occupation, Age, Sex, and Race," *Journal of the American Medical Association,* LXXXV (1925), 1958–1963.

APPENDIX: DATA SOURCES AND CONSTRUCTION

There are two major empirical components of the present study, both of which involve combining micro and aggregate data. The first component is an analysis of sequential cross sections (SCS) from different points in time. That is, I compare a particular age group in one year to that same age group later on, and analyze changes over time differentially by area on the basis of each area's pretreatment-campaign infection level. The second component is a comparison of outcomes across cohorts. This retrospective cohort (RC) analysis is similarly combined with cross-area comparisons based on pretreatment disease burden. In this appendix, I discuss the micro data employed in first the SCS and then the RC analyses. I later describe the construction of the aggregate data on hookworm and the additional control variables that factor into the SCS and RC analysis.

I. SOURCES AND DEFINITIONS FOR THE MICRO DATA

*I.A. Sequential Cross Sections*

The micro data for the SCS component are samples drawn from the Censuses of 1900, 1910, 1920, 1940, and 1950, accessed through the IPUMS project [Ruggles and Sobek 1997; Accessed May 30, 2003]. The sample consists of native-born whites and blacks in the age range [8,16] in the case of children, and in the age range [25,55] in the case of adults. The age criteria for children serves to select children of school age who are likely not yet old enough to have migrated on their own. The lower age cutoff for adults removes those whose school-years were likely impacted by the RSC in 1920. The outcome variables are defined as follows:

- **School enrollment.** This is an indicator variable for whether the child has attended school at any time during a specified interval preceeding the day of the Census. The length of this interval varies across the Censuses as fol-

lows: 1900, within the past year; 1910 and 1920, since September 1; 1940, since March 1; 1950, since February 1.

- **Full-time school attendance.** This is an indicator variable that is switched on if the child is attending and not working. I consider a child to be working if the Census recorded an occupation for him/her, which corresponds to a nonmissing "occ1950" code less than or equal to 970.
- **Literacy.** This variable is an indicator for the ability to read and/or write.
- **Labor-force participation.** A binary variable indicating whether the individual is working. Prior to 1940, this variable is based on whether the individual's reported occupation was classified as a "gainful" one. From 1940 on, the question corresponds more closely to the modern definition.
- **Occupational income score.** See below.

### IB. Retrospective/Cohort

The micro data for the RC analysis is drawn from the IPUMS data. The sample definitions and data construction for Sections VI.A and VI.B are distinct and thus discussed separately.

**Construction of the sample for Table V.** The sample used in Section VI.A consists of native-born whites and blacks in the age range [25,60] in the 1940 Census microdata, except for the literacy sample, which contains native-born whites and blacks with ages [15,45] from the 1920. (The data were accessed February 5, 2003.) The outcome variables are defined as follows:

- **Earnings.** The Census earnings variable from 1940 measures the individual's wage and salary income from 1939. This measure excludes earnings from self-employment.
- **Years of schooling.** I recode the IPUMS "higrade" variable as follows: (i) kindergarten and below to zero and (ii) the remaining values to be the number of years, starting with first grade.
- **Literacy.** Defined above.

**Construction of the sample for Figure III.** The underlying sample used in Section VI.B consists of native-born whites in the age range [25,60] in the 1900–1990 IPUMS microdata and the 1880 microdata from the North Atlantic Population Project [NAPP, 2004]. (These data were last accessed November 14, 2005.) This results in a data set with year-of-birth cohorts from

1825 to 1965. The original micro-level variables are defined as follows:

- **Occupational income score.** The occupational income score is an indicator of income by disaggregated occupational categories. It was calibrated using data from the 1950 Census, and is the average by occupation of all reported labor earnings. See Ruggles and Sobek [1997] for further details.
- **Duncan socio-economic index.** This measure is a weighted average of earnings and education among males within each occupation. The weights are based on analysis by Duncan [1961] who regressed a measure of perceived prestige of several occupations on its average income and education. This measure serves to proxy for both the income and skill requirements in each occupation. It was similarly calibrated using data from the 1950 Census.

These data are used to construct a panel of income by year of birth and state of birth. The cohort-level outcomes are constructed as follows.

1. The microdata from 1880–1990 are first pooled together.
2. The individual income proxies are projected on to dummies for year-of-birth × Census year, i.e. I run the following regression:

$$y_{itk} = \delta_{tk} + \varepsilon_{itk}$$

for individual $i$ in cohort $k$ when observed in census year $t$. This regression absorbs all cohort, age, and period effects that are common for the whole country.

3. I then define cells for each combination of year of birth and state of birth. Within each cell, I compute the average of the estimated income residuals (the $t_{ijk}$). Because these averages are constructed with differing degrees of precision, I also compute the square root of the cell sizes to use as weights when estimating (4).

4. I do this separately for both the occupational income score and the Duncan socio-economic index.

These average income proxies by cohort form the dependent variables in Section VI.B and specifically Figure III.

For the majority of the years of birth, I can compute average income proxies for all of the fifty-one states plus the District of Columbia. The availability of state-level hookworm data and the control variables restricts the sample further to forty-six states of

birth. Hawaii is excluded because of missing data on hookworm. Alaska, Colorado, the District of Colombia and Oklahoma are excluded because of missing data for at least one of the other dependent variables. This leaves forty-six states of birth in the base sample.

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

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

## II. Sources and Definitions for the Aggregate Data

### II.A. Variables Used in the SCS Analysis

Because county boundaries change over time and because county of residence is not available in the later Censuses. I use the state economic area (SEA) as the aggregate unit for the sequential-cross-section analysis, such as in Section. The SEAs are aggregations of counties, with an average number of 8.5

counties per SEA. SEA boundaries tend to be more stable, in part because they were often defined by a state boundary or significant natural feature (river or mountain range, e.g.). (See Bogue [1951] for more detail.)

The area-level data come from a variety of county-level sources, but principally from the RSC annual reports and the ICPSR's study #3, the latter of which is a collection of historical Census tabulations. When relevant, the formulae for constructing the variable are presented below. (Variable names are those of the ICPSR study.) Data refer to 1910 unless otherwise noted. To construct SEA-level data, I sum the constituent counties or construct population-weighted averaged, as appropriate. "Per capita" normalizations come from the ICPSR study #3. The following is a list (in thematic groupings) of the aggregate variables with information on sources and definitions. The method of aggregation is noted if different from above. The source is indicated in brackets at the end of each item.

**Data on Hookworm and RSC Treatments:**

- **Hookworm infection rate.** The source data are at the county level and from the period 1911–1915. The infection numbers in most cases are from surveys conducted by the Rockefeller Sanitary Commission (RSC) as prelude to (or simultaneously with) dispensing treatments. In a few instances, the RSC dispensaries had already visited the county before making the survey. For this latter case. I use the examinations conducted by the dispensaries to construct the hookworm infection rate, rather than using surveys collected after the administration of the RSC treatments. (The hookworm-infection rates constructed from survey and examination have a correlation coefficient greater than 0.95 for those cases in which the survey was done first.) [RSC Annual Reports, 1910–1915.]

- **Individuals treated at least once by the RSC, per capita.** The source data are at the county level and from the period 1911-1915. The RSC dispensaries tracked how many individuals received deworming treatments. If an RSC dispensary visited a county twice, I sum the individuals treated from each visit. While it is possible that some children were double counted in this procedure, generally multiple visits by dispensaries were to cover different territory. [RSC Annual Reports, 1910–1915.]

**Health and Health Policy:**

- **Examined by RSC per capita.** The source data are at the county level and from the period 1911–1915. The RSC tracked how many individuals were examined by the dispensaries' medical staff. [RSC Annual Reports 1910–1915.]

- **Sanitary Index.** The RSC conducted independent surveys of the condition of sanitation infrastructure, including whether buildings had proper latrines, clean water sources, etc. Several measures of sanitation were combined by the RSC to form an index. [RSC Annual Reports 1910–1915.]

- **Full-time health officer.** These data were compiled at the county level, and include information on the first year each county employed a full time health officer. I coded this variable as one if such an office was created between 1910 and 1920 (inclusive). [Ferrell et al. 1932.]

- **County spending.** Data were input at the county level on county-government spending on education and health/sanitation for the years 1902 and 1932. (The 1922 publication in the series did not include these categories of spending, and the 1913 publication did not include earmarked transfers from the state government.) The health spending is normalized by total population, while the education expenditure is normalized by school-age population. [U. S. Bureau of the Census, 1915b and 1935.]

- **WWI cantonment size per capita.** I collected data on the troop numbers that were mustered and trained at the major Army cantonments of mobilization/embarkation for the First World War. Of the thirty-two cantonments, there were nineteen camps in the South. I input the highest value given for the number of soldiers within a camp during 1918–1920. [Bowen 1928.]

- **Malaria mortality, 1919–1921.** [Maxcy 1923.]

- **Change in fertility, 1900–1910.** The fertility rate for 1910 is measured from Census tabulations under the fraction of the population under six years of age, defined as $1 - (v41 + v53)/(v20 + v21)$. For 1900, the tabulations permit calculating the fraction of the population under five for 1900, or $1 - (v22 + v37 + v39 + v41 + v43)/(v8 + v10)$. When computing the approximate difference, I up-weight the 1900 number by 5/4. [ICPSR #3.]

**Education:**

- **Log Change in School Term Length, c. 1905–1925.** Average length of school term, in weeks. Kentucky county data are imputed from cross-tabulated data on number of schools by months. The imputation is calibrated using Alabama data, which contain a continuous measure and a cross-tabulation. [Annual and biennial reports of the various state departments of education, 1905–1930.]
- **Log Change in Average Monthly Salaries for Teachers, c. 1905–1925.** Generally these data were reported directly, but in a few cases, I had to construct the variable using annual salaries and term length. No adjustment for full-time equivalence was available from the source data. [Annual and biennial reports of the various state departments of education, 1905–1930.]
- **Log Change in School Density, c. 1905–1925.** Number of schoolhouse operating in the county, divided by land area in square miles. [Annual and biennial reports of the various state departments of education, 1905–1930; ICPSR #3.]
- **Log Change in Number of Teachers per School, c. 1905–1925.** [Annual and biennial reports of the various state departments of education, 1905–1930.]
- **Log Change in Pupil/Teacher Ratio, c. 1905–1925.** Average attendance divided by number of teachers. [Annual and biennial reports of the various state departments of education, 1905–1930.]
- **Log Change in Value of School Plant and Equipment, c. 1905–1925.** [Annual and biennial reports of the various state departments of education, 1905–1930.]
- **Log Change in County Spending, c. 1905–1925.** See description above with the health controls.
- **Change in Returns to Literacy for Adults, c. 1910–1920.** Measured from a regression of the occupational income score on literacy status, by SEA, for the 1910 and 1920 census samples of adults. (Author's calculations using the 1910 and 1920 IPUMS data.)
- **Literacy Rates.** These data were compiled at the county level and come from the 1910 Census. Child literacy refers to ages 10–20 and is constructed as follows: $1 - (v50/v49)$. Adult literacy refers to males of voting age, defined as $1 - (v37/v26)$. [ICPSR #3.]

**Race and Race Relations:**

- **Fraction black.** These data come from the 1910 Census. Defined as the fraction of the areas males who are black, out of the total population of blacks and whites. Specifically this is defined as $(v24 + v25)/(v24 + v25 + v22 + v23)$. [ICPSR #3.]

- **Rosenwald schools per capita.** This measures the number of classrooms per capita built by the Julius Rosenwald Fund as of 1930. The denominator normalizes the number of classrooms by the population of blacks aged 5–19 in 1930. [Johnson 1941.]

- **Lynchings per capita, 1900–30.** The base data are the number of lynchings per 100,000 population by county in the years 1900–30. The denominator is the county population in 1930. [Johnson 1941.]

**Agricultural/Rural Controls:**

- **Population urban, 1900 and 1910.** From Census tabulations measuring the population residing in metro areas. For 1910, the urban population is contained in variable $v9$ in the ICPSR data, which I scale by the total population as defined above. The 1900 fraction urban is also defined in the 1910 data as $v13/(v13 + v14)$. I construct the change in urbanization using the difference between the two variables. [ICPSR #3]

- **Crop acreage per capita.** The base data measures the total farmed acreage at the county level, regardless of tenancy. This is constructed with the formula $(v155 + v164 + v175)$ and scaled by total population. [ICPSR #3]

- **Sharecropped areas per capita.** The base data are a county-level measure of total acreage sharecropped ($v164$ using the ICPSR variable scheme). I scale this by total population. [ICPSR #3]

- **Farm value per capita.** The base data are a county-level measure of the value of farm land and buildings, regardless of tenancy. This is defined as $(v177 + v166 + v157)$. [ICPSR #3]

- **Cotton acreage per capita.** The base data are cotton acreage in 1910 by county. [Census 1915.]

- **Tobacco acreage per capita.** The base data are tobacco acreage in 1910 by county. [Census 1915.]

**Parental-Background Controls:** The mother's and father's Occupational Income Score are used as indicators for socioeconomic

status. These data are matched to children using the "momloc" and "poploc" variables in the IPUMS. I also construct dummies for parent missing, and assigned them incomes of zero. These variables are interacted with census year in the regressions.

### II.B. Variables Used in the RC Analysis

For the retrospective-cohort analysis, I focus on state on birth as birthplace is not available at further disaggregation. (The District of Columbia is included, where data are available.)

- **Hookworm infection.** Computed from examinations of army recruits. [Koford and Tucker 1921]
- **Average wage, 1899.** I input the average monthly earnings (with board) for farm laborers by state in 1899. Various other wage measures are summarized by the same source, but are generally not available for a complete set of states. [Lebergott 1964; Table A-24]
- **Region of birth.** These dummy variables correspond to the Census definition of regions: Northeast, South, Midwest, and West.
- **Doctors per capita, 1898.** [Abbott 1900.]
- **State public-health spending, 1898.** Per capita appropriations, by state, for state boards of health in 1898. [Abbott 1900.]
- **Child mortality, 1890.** The estimates of child mortality are constructed from published tabulations. Table 3 in Part III contains enumerated deaths of children under one year. I scale this number by the estimated birth rate (Part 1, page 482) times the female population (Part 1, Table 2). [Census, 1894.]
- **Recruits for World War I found rejected for military service because of health "defects," 1917–1919.** Fraction of total record. [Love and Davenport 1920.]
- **Mortality from other diseases.** Separate variables are constructed for the following eight causes of death: scarlet fever, measles, whooping cough, diphtheria/croup, typhoid fever, malaria, diarrheal diseases, and pneumonia. Data are expressed as the fraction of total mortality in 1890. [Census, 1894.]
- **Fertility rate, 1890.** The estimated birth rate (from Part 1, page 182). [Census, 1894.]

- **Log Change in School Term Length, c. 1902–1932.** Average length of school term, in weeks. [Annual reports of the federal Commissioner of Education, 1905–1932.]
- **Log Change in Average Monthly Salaries for Teachers, c. 1902–1932.** [Annual reports of the federal Commissioner of Education, 1905–1932.]
- **Log Change in Pupil/Teacher Ratio, c. 1902–1932.** Average attendance divided by number of teachers. [Annual reports of the federal Commissioner of Education, 1905–1932.]
- **Log Change in School Expenditure, c. 1902–1932.** [Annual reports of the federal Commissioner of Education, 1905–1932.]
- **Adult literacy rate.** Defined as above.
- **Population urban.** Defined as above.
- **Fraction black.** Defined as above.
- **Male unemployment rate, 1930.** [ICSPR #3.]

### References for Data Appendix

Bogue, Donald J., "State Economic Areas: a Description of the Procedure Used in Making a Functional Grouping of the Counties of the United States," Washington, DC: U.S. Government Printing Office, 1951.

Bowen, Albert S., "Activities Concerning Mobilization Camps and Ports of Embarkation. Vol. IV." Washington, DC: U.S. Government Printing Office. Prepared under the direction of The Surgeon General, M. W. Ireland, 1928.

Department of Education of the State of Alabama. *Annual Report.* Montgomery, AL: State of Alabama, 1905–1925.

Department of Education of the State of Georgia. *Annual Report.* Atlanta, GA: State of Georgia. 1905–1925.

Inter-university Consortium for Political and Social Research. Historical, Demographic, Economic, and Social Data: the United States, 1790–1970, Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR). Computer file, 1984 http://www.icpsr.org/. [Accessed Sept. 20, 2002.].

Johnson, Charles S., "Statistical Atlas of Southern Counties: Listing and Analysis of Socio-Economic Indices of 1101 Southern Counties," Chapel Hill, NC: The University of North Carolina Press, 1941.

Love, Albert C., Charles B. Davenport, "Defects Found in Drafted Men: Statistical Information Compiled from the Draft Records Showing the Physical Condition of the Men Registered and Examined in Pursuance of the Requirement of the Selective Service Act," Washington, DC: U.S. Government Printing Office. Prepared under the direction of The Surgeon General, M.W. Ireland, 1920.

Maxcy, Kenneth F., "The Distribution of Malaria in the United States as Indicated by Mortality Reports," Public Health Reports, XXXVIII (1923), 1125–38.

Superintendent of Education of the State of South Carolina. *Annual Report.* Columbia, SC: State of South Carolina. 1905–1925.

Superintendent of Public Education of the State of Louisiana, *Biennial Report.* Baton Rouge, LA: State of Louisiana. 1905–1925.

Superintendent of Public Education of the State of Mississippi, *Biennial Report.* Jackson, MS: State of Mississippi. 1905–1925.

Superintendent of Public Instruction of the Commonwealth of Kentucky, *Biennial Report.* Frankfort, KY: Commonwealth of Kentucky, 1905–1925.

Superintendent of Public Instruction of the Commonwealth of Virginia, *Annual Report.* Richmond, VA: Superintendent of Public Printing. 1905–1925.

Superintendent of Public Instruction of the State of Arkansas, *Biennial Report.* Little Rock, AR: State of Arkansas. 1905–1925.

Superintendent of Public Instruction of the State of North Carolina, *Biennial Report.* Raleigh, NC: State of North Carolina. 1905–1925.

Superintendent of Public Instruction of the State of Tennessee, *Biennial Report.* Nashville, TN: State of Tennessee. 1905–1925.

Superintendent of Public Instruction of the State of Texas, *Biennial Report.* Austin, TX: State of Texas. 1905–1925.

United States Bureau of the Census, "Vital and Social Statistics," *Compendium of the Eleventh Census, 1890,* (Vol. II), Washington, DC: U.S. Government Printing Office, 1894.

——, "Report on the Statistics of Agriculture in the United States," *Thirteenth Decennial Census of the United States,* 1910. (Vol. VI & VII). Washington, DC: United States Government Printing Office, 1915a.

——, *Wealth, Debt and Taxation.* Vol. II. Washington, DC: United States Government Printing Office, 1915b.

——, *Financial Statistics of State and Local Governments: 1932.* Washington, DC: United States Government Printing Office, 1935.

United States Office [Bureau] of Education, *Annual Report of the Commissioner Education.* Washington, DC: Government Printing Office, 1894–1913.