



CHAPTER 5

Disease and Development: Evidence from Hookworm Eradication in the American South

Hoyt Bleakley

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.

Reprinted with permission from Hoyt Bleakley, "Disease and Development: Evidence from Hookworm Eradication in the American South," *The Quarterly Journal of Economics* (The MIT Press) 122, no. 1 (February 2007): 73–117. © 2007 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

This is a revised version of the first chapter of the author's doctoral dissertation (Bleakley 2002b), which was partially summarized in Bleakley (2003). The author thanks Daron Acemoglu, Joshua Angrist, David Autor, Gary Becker, Eli Berman, Patrick Buckley, Garland Brinkley, Dora Costa, Mark Duggan, Michael Greenstone, Jonathan Guryan, Christian Hansen, Gordon Hanson, Lakshmi Iyer, Simon Johnson, Lawrence Katz, Fabian Lange, Mark Lewis, Robin McKnight, Derek Neal, John Strauss, Robert Triest, Burton Weisbrod, Jonathan Zinman, several anonymous referees, and seminar participants at Boston University, the University of Chicago, Harvard University, Illinois State University, the Massachusetts Institute of Technology, Northwestern University, Princeton University, the University of California, Berkeley, University of California, San Diego, the University of Southern California, and Yale University for useful comments, and Michael Pisa, Tareq Rashidi, and Elizabeth Stone for excellent research assistance.

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

This chapter focuses on one specific intervention targeted toward hookworm disease in the American South. The hookworm eradication campaign (circa 1910–15) began soon after (a) the discovery that a variety of health problems among Southerners could be attributed to the disease and (b) 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-age 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 neither related to nor 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, 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 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 rule out the possibility that the intervention had effects through channels besides measured hookworm infection. Replicating this design using state-of-birth-level 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. However, 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 (a) the pre-eradication hookworm burden in their state of birth and (b) 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 caution as above), I estimate that being infected with hookworm throughout 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 coincided 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 chapter is organized as follows. The first section describes the symptoms and history of the disease, in particular how the circumstances of the discovery of the hookworm problem in the South and the subsequent antihookworm campaign lend themselves to a strategy for identifying the effect of hookworm. The second section describes the data employed, the third and fourth present the contemporaneous results using sequential cross sections, and the fifth section presents the long-term follow-up. The last section summarizes the findings and suggests areas of future study.

Hookworm and the Rockefeller Sanitary Commission

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 (*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 (for example, the use of sanitary latrines).

The Eradication Campaign

The Rockefeller Sanitary Commission for the Eradication of Hookworm Disease was formed in 1910 with the donation of \$1 million 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. It surveyed more than 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 more than 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 dropped by more than half, and fewer extreme cases of the disease went unnoticed and untreated.

1 Thymol, taken orally, was the recommended treatment of the time.

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 government actors likely played a key role in consolidating the gains from the RSC campaign and continuing the progress toward complete eradication.² State governments ramped up their funding of antihookworm 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 entered into the public consciousness.

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 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

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 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 few long-term benefits. This chapter 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.

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.

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 residents of 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 eradication campaign was made possible by critical innovations to knowledge: understanding how the disease worked and, more important, 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.

Third, the antihookworm 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 50-year time span, which is unquestionably long relative to the five-year RSC intervention.

3 The historical presentation in this section draws heavily on the work of Ettlting (1981).

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

$$(\text{Pretreatment Infection Rate})_j \times (\text{Indicator for Posttreatment})_t.$$

More compactly, call this variable $H_j^{\text{pre}} \times \text{Post}_t$, where j indexes the geographic area and t indicates the year. The variable H_j^{pre} denotes the level of hookworm infection among school-age children in area j at the time of the RSC's initial survey, and Post_t is a dummy variable indicating whether year t is later than the active years of the RSC campaign (1910–15).

I compare the evolution of outcomes (such as 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 equation 5.1 measures the reduced-form differences by pre-eradication hookworm for some outcome Y_{ijt} for person i in area j at time t .⁴

$$Y_{ijt} = \beta(H_j^{\text{pre}} \times \text{Post}_t) + \delta_t + \delta_j + X_{ijt}\Gamma + \varepsilon_{ijt}, \quad (5.1)$$

in which Y_{ijt} is the outcome of interest, δ_t are time dummies, δ_j are geographic fixed effects, and X_{ijt} is some vector of individual-level controls.⁵

How realistic is the assumption that areas with high infection rates benefited more from the eradication campaign? Follow-up surveys 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 5.1 presents data on this issue.⁶ The basic assumption here—that areas where hookworm was highly endemic saw a greater drop in infection than areas with low infection rates—is borne out across states and across counties.

4 All of the estimates of this equation are calculated using ordinary least squares 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:

$$Y_{ijt} = \alpha H_{ijt} + \delta_j + \delta_t + X_{ijt}\Gamma + \bar{\varepsilon}_{ijt},$$

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 (that is, aggregate) counterpart:

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

(This equation can equally be run in aggregate form entirely, and, when estimated, it gives very similar results to those found here.) For the instrument $H_j^{\text{pre}} \times \text{Post}_t$, the reduced form of this system is equation 5.1. 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:

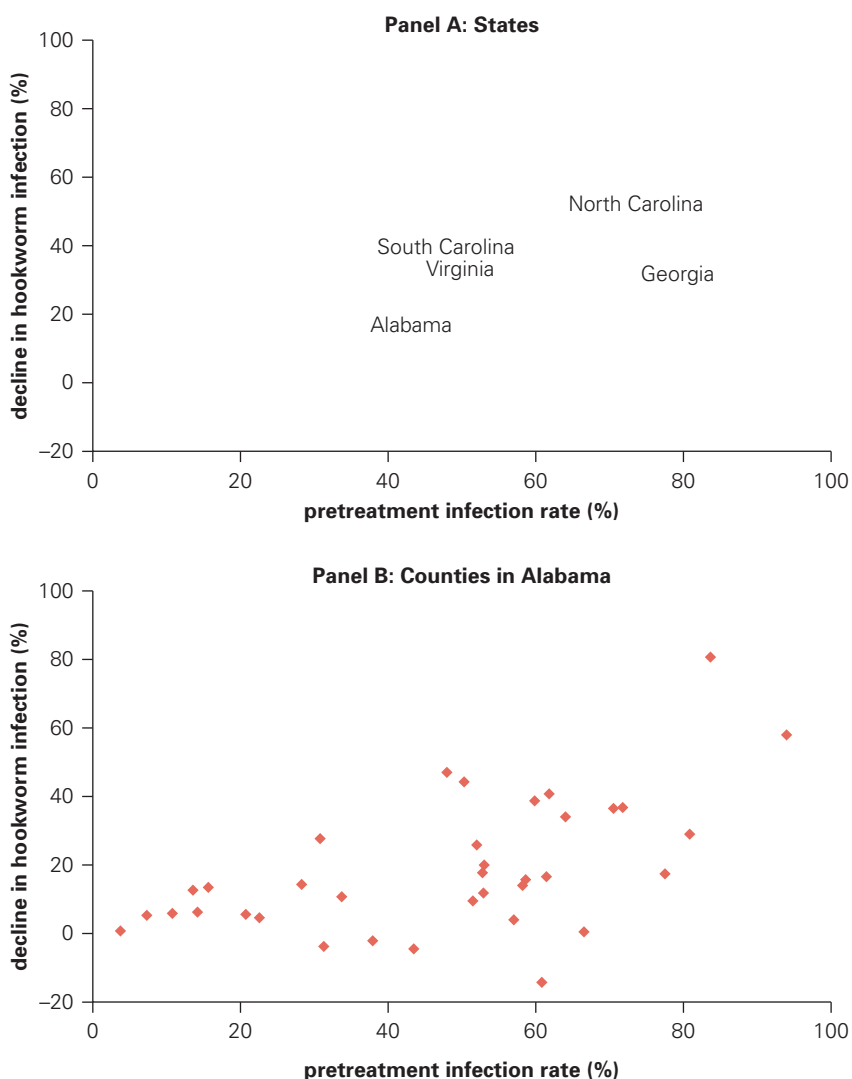
$$H_{jt} = \gamma(H_j^{\text{pre}} \times \text{Post}_t) + \delta_j + \delta_t + \eta_{jt}.$$

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

$$\Delta H_{jt}^{\text{post}} = \gamma H_j^{\text{pre}} + \text{constant} + v_{jt},$$

an equation that relates the observable variables graphed in figure 5.1.

Figure 5.1 Hookworm Infection Rates Before and After Treatment at the State Level in Select Southern States and at the County Level in Alabama



Note: The y axis displays the decrease in hookworm infection after intervention, 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 follow-up surveys are from the early 1920s. The average number of children examined per country exceeds 450 in both studies.

Related Studies

Several pieces of contemporaneous evidence also complement the results from this study. Summarizing evidence from randomized trials in developing countries, Dickson and others (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 failed to deal with the reinfection problem. As a result, they argue, follow-up surveys often found limited effects; no increase in school attendance was observed because there was little persistent difference in infection rates between control and treatment 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 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.

Several other recent studies consider the early-twentieth-century reduction in tropical diseases in the American South. While childhood effects are the focus of this chapter, Brinkley (1994) examines the role that 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.

Data and Descriptive Statistics

This study links aggregate data on hookworm infection with individual socioeconomic data. Table 5.1 contains summary statistics of various outcomes. Because county boundaries change, I use aggregated county groupings, or “state economic areas” (SEAs), as the geographic unit.

The hookworm infection rates were computed by the Rockefeller Sanitary Commission for more than 550 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-age children in each county.⁷ The RSC surveys measured an unweighted average infection rate across SEAs of 32 percent.

The number of individuals treated at least once by the RSC (scaled by 1910 SEA youth population) is reported in table 5.1. 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 or herself at a commission dispensary, the RSC directed more resources toward the areas with greater

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 6 and 18 taken at random—that is, without reference to clinical symptoms—from rural districts distributed over the county.”

Table 5.1 Summary Statistics

Indicator	Whole sample	By hookworm infection		Source
		>40%	<40%	
Hookworm infection rate	0.320 (0.230)	0.554 (0.137)	0.164 (0.117)	RSC annual reports
Individuals treated at least once by the RSC, per school-age child	0.206 (0.205)	0.342 (0.199)	0.109 (0.147)	RSC annual reports
School enrollment, 1910	0.721 (0.104)	0.711 (0.099)	0.729 (0.108)	IPUMS; author's calculations
Change in school enrollment, 1910–20	0.089 (0.080)	0.103 (0.090)	0.078 (0.072)	IPUMS; author's calculations
Full-time school attendance, 1910	0.517 (0.140)	0.469 (0.123)	0.551 (0.141)	IPUMS; author's calculations
Change in full-time school attendance, 1910–20	0.203 (0.097)	0.246 (0.093)	0.172 (0.089)	IPUMS; author's calculations
Literacy, 1910	0.853 (0.104)	0.824 (0.101)	0.875 (0.102)	IPUMS; author's calculations
Change in literacy, 1910–20	0.060 (0.067)	0.081 (0.075)	0.045 (0.057)	IPUMS; author's calculations
Share of the population that is black, 1910	0.357 (0.221)	0.41 (0.208)	0.318 (0.223)	IPUMS; author's calculations
Fraction of the population that is urban, 1910	0.174 (0.200)	0.167 (0.214)	0.180 (0.190)	ICPSR (1984)
School term, in months, c1910	5.251 (1.066)	5.055 (1.042)	5.391 (1.068)	State annual reports
Schools per square mile, c1910	0.195 (0.358)	0.142 (0.053)	0.233 (0.465)	State annual reports; ICPSR
Value of school property, per pupil, current dollars, c1910	5.518 (4.037)	4.699 (3.159)	6.104 (4.496)	State annual reports
Teacher-to-school ratio, c1910	1.336 (0.545)	1.397 (0.505)	1.293 (0.572)	State annual reports
Sample size	115	48	67	n.a.

Source: Author's calculations.

Note: Variable means are displayed in the first column. Standard deviations are displayed in parentheses below the mean. Sample selection for the Integrated Public Use Micro Sample (IPUMS) data consists of native-born whites and blacks, in the RSC-surveyed geographic units, for the indicated years. The school enrollment and attendance data are constructed for children ages 8–16; literacy data are for children ages 10–16, and the RSC reported infection rates are for children ages 8–16. See the appendix for further information on sources and construction of the variables. n.a. = not applicable.

hookworm infection. These numbers indicate that about 64 percent of the infected population received deworming treatments.⁸

The micro-level data come from the IPUMS, a project harmonizing the coding of historical census micro data (Ruggles and Sobek 1997). The RSC's

⁸ Similarly, a regression with these measures at the SEA level ($N = 113$; $R^2 = 0.495$) yields the following estimates:

$$T_{ij} = 0.619H_i^{\text{pre}} + 0.003 + \varepsilon_{ij},$$

(0.064) (0.017)

where T_{ij} is the number of individuals treated at least once by the RSC, divided by the school-age population (ages 6 to 17, inclusive) in SEA_{*i*}. This indicates that, on the margin, about 62 percent of sufferers were treated.

activities took place from 1910 to 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 to 1950, and the long-term follow-up comprises census samples from 1880 to 1990.

Three binary indicators of human capital are used here: 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 are coded as attending school regularly if they were enrolled in school and they did not report an occupation. The literacy variable indicates whether the child could 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 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 hookworm-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 the prevalence of one-room common schools. These variables and others constitute important controls in the sensitivity analysis below.¹⁰

Contemporaneous Effects on Children

Here I conduct regression analyses of changes in literacy, school enrollment, and school attendance between the 1910 and 1920 censuses by estimating equation 5.1. 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 $H_t^{pre} \times Post_t$ 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

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.

10 In previous versions of this study, I also compare hookworm infection with the prevalence of other disease conditions. I find a relationship between hookworm and malaria across the counties of the South. However, 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.

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 school enrollment of one-quarter of a standard deviation.

These results are presented in table 5.2. Estimates of the variable of interest— $H_j^{\text{pre}} \times \text{Post}_t$ —are displayed for various outcomes and specifications. Panel A presents the main results. The first row contains the estimates using the 1910 and 1920 censuses, which bracket the RSC intervention, while the second row contains similar estimates using the census micro data from 1900 to 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 blank in the first and second rows of panel A; literacy results in panel B use the 1910–20 censuses.)

Table 5.2 Hookworm and Human Capital: Basic Results

Dependent variables	Estimating equation	School enrollment (1)	Full-time school attendance (2)	Literacy (3)
<i>Panel A: Basic results</i>				
Census years				
1910–20	5.1	0.0883*** (0.0225)	0.1591*** (0.0252)	0.0587*** (0.0186)
1900–50	5.1	0.0608*** (0.0261)	0.1247*** (0.0286)	
1900–50	5.2	0.0954*** (0.0233)	0.1471*** (0.0287)	
<i>Panel B: Effects within and between states</i>				
Change to specification				
Include state × Post dummies		0.1313*** (0.0245)	0.2144*** (0.0290)	0.0417** (0.0207)
Allow for state-specific mean reversion		0.1148*** (0.0265)	0.1813*** (0.0312)	0.0408** (0.0206)
Use infection from state of birth instead of SEA		0.0489 (0.0504)	0.2057*** (0.0765)	0.0907** (0.0451)
Census years		1900–50	1900–50	1910–20
Estimating equation		5.2	5.2	5.1

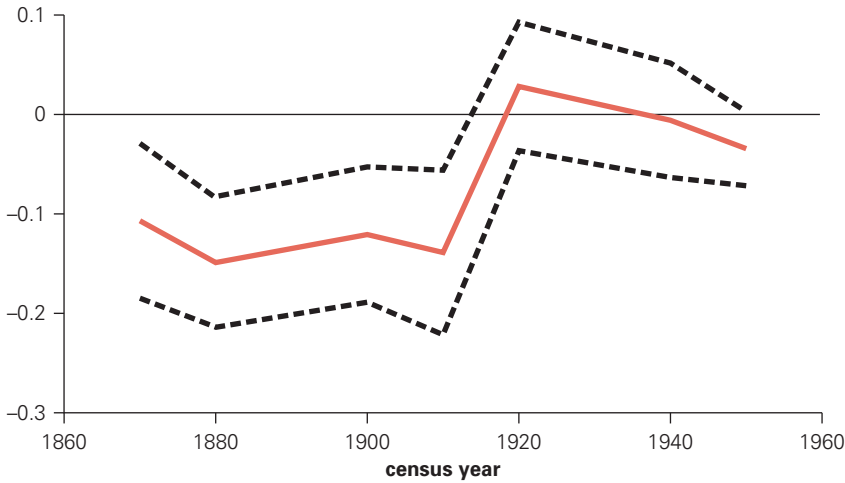
Source: Author's calculations.

Note: 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 are in parentheses (clustering on area × Post_t). 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 the second and third rows of panel B. 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 panel A and the first two rows of panel B, the sample is drawn from the RSC-surveyed county groups (SEAs), all of which were in the South. For the third row of panel B, 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 the first and second rows of panel A.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

Figure 5.2 Hookworm Eradication and School Attendance, 1870–1950



Source: Author's calculations.

Note: 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 1870, 1880, 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.

The surge in school attendance in high-hookworm counties coincided with the campaign for hookworm eradication. This can be seen in figure 5.2.¹¹ 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 equation 5.1, resulting in the following equation:

$$Y_{ijt} = \beta(H_j^{\text{pre}} \times \text{Post}_t) + \tilde{\delta}_j \times t + \delta_t + \delta_j + X_{ijt}\Gamma + \varepsilon_{ijt}. \quad (5.2)$$

Trend differences across areas load onto the $\tilde{\delta}_j$, while differences that coincide with the antihookworm campaign load onto β . Estimates of equation 5.2 in row C of table 5.2 show little change in the estimated β .

The specification in the first row of panel B 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 \times year, this specification implements a simple fix to purge the estimates of this effect, including (state \times 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 school attendance to rise in the following period even if hookworm had not affected the schooling decision. In the third row of

¹¹ To construct this figure, I run a regression of school attendance on SEA-level hookworm, separately by census year from 1870 to 1950. Micro-level controls for age, female, female \times age, black, and black \times age are also included. The year-specific estimates on H_j^{pre} are plotted against year.

panel B of table 5.2, I add the interaction of $Post_t$ with 1910 average school attendance by SEA. Differential incidence of state policies (by average school attendance rates) are also absorbed by interacting state \times 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 the third row of panel B, I reestimate equation 5.2 using only state-of-birth-level variation in the antihookworm 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: (a) there are fewer geographic units, and (b) the dispersion of infection rates across states is smaller than that across county groups.

However, 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. In contrast, 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

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 equations 5.1 and 5.2 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 5.3. 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 5.3. For preteens and adolescents, the estimates for enrollment 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 below, since the retrospective-cohort analysis employs precisely the state-of-birth variation in hookworm.

Table 5.3 Hookworm and Human Capital: Sensitivity Tests and Results for Subgroups

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

Source: Author's calculations.

Note: This table reports estimates of the interaction of pretreatment hookworm and a post-RSC dummy in equation 5.1 (for the 1910–20 data) and equation 5.2 (for the 1900–50 data), for the indicated subsamples. The dependent variables are the binary indicators denoted in the column headings. Robust standard errors are in parentheses (clustering on SEA × Post). The 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. The number of clusters is 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. Reporting of additional coefficient estimates is suppressed.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

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 antihookworm campaign. Whites appeared to have positive responses to hookworm eradication by all three measures of human capital, but the estimated effects of eradication were uniformly larger for blacks. There are several possible 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. The other possibility is that whites, because of higher average incomes and 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 for improvement along these measures of human capital investment.¹³ The long-term consequence of these racial differences is less clear because the return to schooling was lower for blacks than for whites during this period. I revisit this issue in a later section.

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 $H_t^{pre} \times Post_t$ (about 0.09) to the estimated decline in infection as a function of the same variable (0.44). The latter number comes from the follow-up surveys discussed above and is shown in figure 5.1. Some of this relationship may be due to “Galton’s fallacy” because it is a comparison of ΔH_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 However, the intervention may have lowered 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, 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 (that is, 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 computing the depressing effect of hookworm on the region’s accumulation of human capital, 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.

13 Using logit and probit estimators rather than a linear probability model, I find that the hookworm effect was sometimes larger for whites, sometimes not, depending on the specification. This bolsters the hypothesis that the two groups were experiencing similar increases in some latent measure (human capital investment), but the 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 above, 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 used in this chapter. However, the estimates are not directly comparable because Miguel and Kremer use a combined infection rate that includes hookworm, roundworm, schistosomiasis, and whipworm.

Contemporaneous Effects on Adults

Next I examine how adult outcomes in the same time periods respond to the antihookworm 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 5.4, which contains estimates of equation 5.1. For several outcomes, I cannot reject the null hypothesis that there was no differential change across counties 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; neither do I find significant effects when I perform the analysis separately by broad age groupings.

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

Table 5.4 Contemporaneous Effect on Adult Outcomes

parameter estimates

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

Source: Author's calculations.

Note: Each cell reports the coefficient estimate on Hookworm \times Post for the indicated sample and dependent variable. Robust standard errors are in parentheses (clustering on SEA \times Post; the number of clusters is 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 1910–20 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.

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.

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 appear to extend into adulthood for the affected cohorts. This section contains several results supporting this conclusion.

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 antihookworm campaign times the pre-eradication hookworm intensity in the state of birth. Exposure to the RSC, Exp_{ik} , is zero for older cohorts, rises linearly for those born in the 19 years prior to 1910, and stops at 19 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 as follows:

$$Y_{ijk} = \beta(H_j^{pre} \times Exp_{ik}) + \delta_j + \delta_k + X_{ijk}\Gamma + v_{ijk} \quad (5.3)$$

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 5.5 contains these results. Panel A presents the estimates of equation 5.3.

The estimates do not appear to be 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 5.5. This analysis yields mixed evidence of mean reversion in the data, but the inclusion of these controls does not substantially affect the coefficient on $H_j^{pre} \times Exp_{ik}$.¹⁸

I argue that the earnings results are not contaminated by hookworm-induced changes in the probability of self-employment. The major difficulty

17 Specifically, the formula is $Exp_{ik} = \max[\min(19, 49 - age_i), 0]$, where age is measured in 1940.

18 Similar results are obtained by interacting the 1899 wage measure with the exposure variable instead of age or including the square of the average wage (\times age) as well.

Table 5.5 Long-Term Follow-up Based on Intensity of Exposure to the Treatment Campaign

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

Source: Author's calculations.

Note: Each panel-column 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 are in parentheses (clustering on state of birth). The demographic controls consist of indicator variables for each age × black × female cell, plus interactions of state-of-birth dummies with black, female, and black × female. Reporting of additional coefficient estimates is suppressed.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

in using the earnings data from the 1940 census is that they are incomplete: labor income from self-employment is excluded. To gauge the impact of this problem, I estimate regression equations identical to those used for panel A of table 5.5, but with three new dependent variables: binary indicators for (a) self-employment, (b) missing data for log earnings, and (c) nonwage or salary income greater than \$50. In doing so, I find no robust and statistically significant relationship between the hookworm measure and any of these three measures.¹⁹ Additionally, I find evidence of a hookworm-related

19 Using controls for mean reversion, the coefficient (standard error) on $H_i \times \text{Exp}_{ik}$ 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 nonwage or salary income exceeding \$50.

increase in the total time worked (either for a wage or a 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 that 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 shown above, students were less likely to work while in school and 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 5.5. The crucial interaction is the triple interaction: 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 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 $H_j \times \text{Exp}_{ik}$ 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 5.5, panel C. For no subgroup is there a robustly significant relationship

20 Similarly, I do not find evidence that the mechanism is migration out of the South, migration into an urban area, shifting to a white-collar occupation, 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 \times education and birth year \times education. The first-order effects of education, state of birth, and year of birth are also absorbed with indicator variables.

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

between years of schooling and $H_j \times \text{Exp}_{ik}$. Estimates for literacy, in contrast, 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 $H_j \times \text{Exp}_{ik}$ 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. In contrast, whites and blacks show similar earnings responses. This may be because blacks, while gaining more on measured human capital, faced lower returns to skill in the labor market.

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 or autoregressive processes.

I use two income proxies that are available for a large number of censuses. The occupational income score and the Duncan socioeconomic index are both average indicators by disaggregated occupational categories that were calibrated using data from the 1950 census. The former variable is the average by occupation of all reported labor earnings. The measure due to Duncan (1961) is instead a weighted average of earnings and education among males within each occupation. Both variables can therefore measure shifts in income that take place between occupations. The Duncan measure has the added benefit of picking up between-occupation shifts in skill requirements for jobs. Occupation has been measured by the census for more than a century, and so these income proxies are available for a substantial stretch of cohorts.

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 10 censuses (1880 and 1900–90). This results in an unbalanced panel spanning the year-of-birth cohorts from 1825 to 1965 for 46 states of birth. (See the appendix for details of the data construction.)

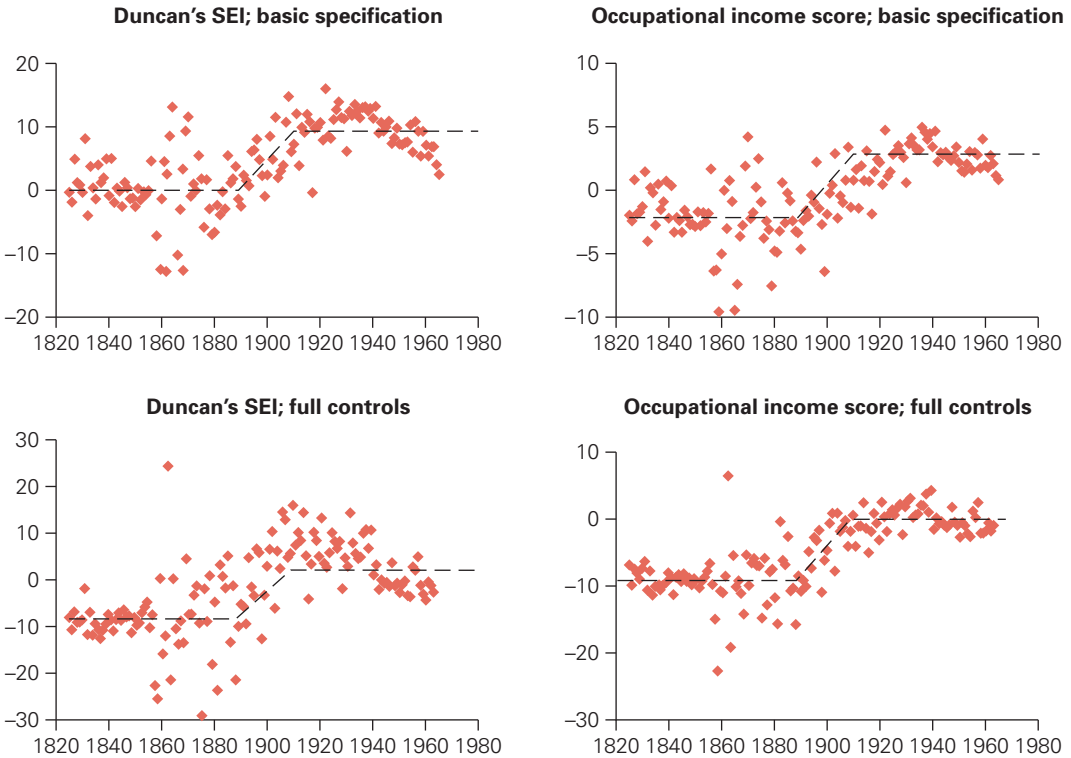
For each year of birth, ordinary least squares (OLS) regression coefficients are estimated on the resulting cross section of states of birth. Consider a simple regression model of an average outcome, Y_{jk} , for a cohort with state of birth j and year of birth k :

$$Y_{jk} = \beta_k H_j^{\text{pre}} \times \delta_k + X_j' \Gamma_k + v_{jk}, \quad (5.4)$$

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

I start with a simple graphic analysis using this flexible specification for cross-cohort comparison. Figure 5.3 displays a plot of the estimated β_k .

Figure 5.3 Cohort-Specific Relationship between Income and Pre-eradication Hookworm



Source: Author's calculations.

Note: This figure summarizes regressions of income proxies on pre-eradication hookworm infection rates by state of birth. 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 underlying 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 control variables described in the appendix. The regressions are estimated using weight equal to the square root of the cell size in the underlying micro data.

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 living in urban areas, 1890 child mortality rate, fraction of deaths in 1890 caused by scarlet fever, measles, whooping cough, diphtheria or 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, World War I recruits found "defective" at draft physical, and 1902–32 logarithmic changes in average monthly teacher salaries, length of school term, school expenditures per capita, and pupil-teacher ratios. See the appendix for details on these data.

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. However, later cohorts experienced reduced hookworm infection during their childhood. This benefit increased with younger cohorts who 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 relationship between income and pre-eradication hookworm 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:

$$\hat{\beta}_k = \alpha \text{Exp}_k + \sum_{i=1}^n \gamma_n k^n + \Phi(L) \hat{\beta}_k = \eta_{jk}, \quad (5.5)$$

in which 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 5.6 reports estimates of equation 5.5 under a variety of order assumptions about trends and autoregression. The dependent variables are $\hat{\beta}_k$, 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 socioeconomic indicator exhibit both trend and autoregression, but the exposure coefficient is stable once both are accounted for.²⁵

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 no childhood exposure to the RSC and the cohort with full

24 The exposure variable is measured in different units than the cohort-specific regression coefficients, so to facilitate visual comparisons in figure 3.3, the line is rescaled in the y dimension so that the pre-1880 and post-1930 levels match those of the β_k . 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 100 and then 40 years of birth cohorts yields similar estimates on exposure, albeit with larger standard errors. I have experimented with higher-order polynomial trends and found no estimates of exposure that are statistically significant for $n \leq 5$.

Table 5.6 Exposure to RSC Versus Alternative Time-Series Relationships*time-series estimates of the exposure coefficient*

Income proxy and specification	(1)	(2)	(3)	(4)	(5)
<i>Occupational income score</i>					
Basic	0.3113*** (0.0214)	0.2915*** (0.0542)	0.2612*** (0.0384)	0.2497*** (0.0612)	0.1912*** (0.0622)
Full controls	0.2623*** (0.0339)	0.3732*** (0.0858)	0.2346*** (0.0438)	0.3393*** (0.0960)	0.2743*** (0.1007)
<i>Duncan's socioeconomic indicator</i>					
Basic	0.5352*** (0.0418)	0.7566*** (0.1069)	0.3928*** (0.0520)	0.5983*** (0.1124)	0.4858*** (0.1282)
Full controls	0.5007*** (0.0661)	0.8820*** (0.1707)	0.3544*** (0.0735)	0.6616*** (0.1791)	0.7081*** (0.1969)
<i>Additional controls</i>					
Order of polynomial trend	0	1	0	1	2
Order of autoregressive process	0	0	1	1	2

Source: Author's calculations.

Note: This table reports the estimates of childhood exposure to the RSC in equation 5.5. The dependent variables are the cohort-specific regression estimates of income proxies on hookworm that are shown in figure 5.3. Robust standard errors are given in parentheses. Observations are weighted by the inverse of the standard error for β_k . In the basic specification, the income proxy was regressed onto hookworm infection, Lebergott's measure of 1899 wage levels, and dummies for the four census regions. The "full controls" specification contains, in addition, the various control variables listed in the appendix.

***Significant at the 1 percent level.

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, 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^{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 the decrease in infection rates as a function of H_j^{pre} to be 0.748.²⁶ This yields an ILS estimate of -0.43 in natural-log terms.

²⁶ This is a departure from the methodology presented in the section on interpretation in that I scale the reduced-form coefficient by the preexisting hookworm infection rate rather than the change. For the ILS calculation, 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 time hookworm had been mostly eradicated. Because eradication was slightly less than complete, this induces a slight downward bias of the ILS estimate. However, 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.

The ILS estimates using the occupational proxies for income are similar to the wage result. The shift in income related to RSC exposure is estimated in table 5.6. 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 the proportional change in income related to childhood hookworm infection to be -0.23 . The same estimate for the Duncan socioeconomic indicator 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 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 in 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 17 log percentage points. In other words, some 22 percent of this income gap could be attributed to hookworm infection in the South. However, 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-age 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 the levels of high-income countries. 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.

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 (a) its timing was relatively short and well defined, (b) geographic differences in infection permit a treatment-control design, and (c) 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? In this chapter, 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 This chapter suggests potentially large benefits of public health interventions in developing countries, where hookworm is still endemic today. Nevertheless, using a simple calculation, I show that, although 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 distribution of income.

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 literature on institutions 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, 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 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.

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.

Appendix: Data Sources and Construction

There are two major empirical components of this chapter, 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 level of infection before the treatment campaign. 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.

Sources and Definitions for the Micro Data

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). 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 affected by the RSC in 1920. The outcome variables are defined as follows:

- *School enrollment*. This is an indicator variable for whether the child attended school at any time during a specified interval preceding the day of the census. The length of this interval varies across the censuses as follows: 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 was attending school and not working. I consider a child to be working if the census recorded an occupation for him or 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, write, or both.
- *Labor force participation*. A binary variable indicating whether the individual was 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.

The micro data for the retrospective cohort analysis are drawn from the IPUMS data. The sample definitions and data construction for sections on the results for earnings, schooling, and literacy and on the cohort-specific relationship between income and hookworm are distinct and thus discussed separately.

The sample for table 5.5 was constructed as follows. The sample used in the section on earnings, schooling, and literacy consists of native-born whites and blacks in the age range [25, 60] in the 1940 census micro data, except for the literacy sample, which consists of native-born whites and blacks with ages [15, 45] from the 1920 census. (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: (a) kindergarten and below to zero and (b) the remaining values to be the number of years, starting with first grade.
- *Literacy*. Defined above.

The sample for figure 5.3 was constructed as follows. The underlying sample used consists of native-born whites in the age range [25, 60] in the 1900–90 IPUMS micro data and the 1880 micro data 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 socioeconomic 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 the occupation's average income and education. This measure serves to proxy for both the income and the 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 1880–90 are first pooled together.
2. The individual income proxies are projected onto dummies for year of birth \times census year, that is, 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 equation 5.4.
4. I do this separately for both the occupational income score and the Duncan socioeconomic index.

These average income proxies by cohort form the dependent variables in the section on cohort-specific relationship between income and hookworm, specifically figure 5.3.

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

Several cohorts were born before 1885 for which as few as 37 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 percent sample for 1880, there are no micro data for 1890, and IPUMS data for 1900 are only a 1 percent sample. In contrast, for the 1843–55 birth cohorts, all but two of the years have all 46 states represented. Nevertheless, even with the 100 percent sample from 1880, there are as many as six states per year missing for those cohorts born before 1843. Several 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 nineteenth century, and it is quite possible that, in certain years, no one eligible to be enumerated was born in some territories. (Untaxed Indians were not counted in the censuses.) Note that I use the term state 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).

Sources and Definitions for the Aggregate Data

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. 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 (such as a river or mountain range). 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 (RSC 1910–15) and ICPSR (1984), which is a collection of historical census tabulations. When relevant, the formulas 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 averages, as appropriate. “Per capita” normalizations come from ICPSR (1984). 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 parentheses at the end of each item.

The following data are used to describe hookworm and RSC treatments:

- *Hookworm infection rate.* The source data are at the county level and from the period 1911–15. The infection numbers in most cases are from surveys conducted by the Rockefeller Sanitary Commission as prelude to (or simultaneously with) dispensing treatments. In a few instances, the RSC dispensaries had already visited the county before conducting the survey. In this case, I use the examinations conducted by the dispensaries to construct the hookworm infection rate, rather than surveys collected after 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 1910–15.)
- *Individuals treated at least once by the RSC per capita.* The source data are at the county level and from the period 1911–15. 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, dispensaries generally made multiple visits to cover different areas. (RSC 1910–15.)

For health and health policy, the following data are used:

- *Number of individuals examined by RSC per capita.* The source data are at the county level and from the period 1911–15. The RSC tracked how many individuals were examined by the dispensaries’ medical staff. (RSC 1910–15.)
- *Sanitary index.* The RSC conducted independent surveys of the condition of sanitation infrastructure, including whether buildings had proper latrines, clean water sources, and so forth. Several measures of sanitation were combined by the RSC to form an index. (RSC 1910–15.)
- *Full-time health officer.* These data are compiled at the county level and include information on the first year each county employed a full-time

health officer. I code this variable as 1 if such an office was created between 1910 and 1920 inclusive. (Ferrell and others 1932.)

- *County spending*. Data are input at the county level on county-government spending on education and health or sanitation for the years 1902 and 1932. (The 1922 publication in the series does not include these categories of spending, and the 1913 publication does 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, 1935.)
- *World War I cantonment size per capita*. I use data on the troop numbers that were mustered and trained at the major Army cantonments of mobilization or embarkation for World War I. Of the 32 cantonments, 19 camps were in the South. I input the highest value given for the number of soldiers within a camp during 1918–20. (Bowen 1928.)
- *Malaria mortality, 1919–21*. (Maxcy 1923.)
- *Change in fertility, 1900–10*. The fertility rate for 1910 is measured from census tabulations under the fraction of the population under six years of age, defined as $1 - (v_{41} + v_{53}) / (v_{20} + v_{21})$. For 1900, the tabulations permit calculating the fraction of the population under five for 1900, or $1 - (v_{22} + v_{37} + v_{39} + v_{41} + v_{43}) / (v_8 + v_{10})$. When computing the approximate difference, I up-weight the 1900 number by five-fourths. (ICPSR 1984.)

For education, the following data are used:

- *Log change in school term length, c1905–25*. This measures the average length of school term, in weeks. Kentucky county data are imputed from cross-tabulated data on number of schools by month. 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, 1905–30.)
- *Log change in average monthly salaries for teachers, c1905–25*. 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, 1905–30.)
- *Log change in school density, c1905–25*. Number of schoolhouses operating in the county, divided by land area in square miles. (Annual and biennial reports of the various state departments of education, 1905–30; ICPSR 1984.)
- *Log change in number of teachers per school, c1905–25*. (Annual and biennial reports of the various state departments of education, 1905–30.)
- *Log change in pupil-teacher ratio, c1905–25*. Average attendance divided by number of teachers. (Annual and biennial reports of the various state departments of education 1905–30.)

- *Log change in value of school plant and equipment, c1905–25.* (Annual and biennial reports of the various state departments of education 1905–30.)
- *Log change in county spending, c1905–25.* See the description above with the health controls.
- *Change in returns to literacy for adults, c1910–20.* This is 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 1984.)

For race and race relations, the following data are used:

- *Fraction black.* These data come from the 1910 census and are defined as the fraction of the area's 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 1984.)
- *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 ages 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.)

For agricultural-rural controls, the following data are used:

- *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 1984.)
- *Crop acreage per capita.* The base data measure 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 1984.)
- *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 1983.)
- *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 1984.)
- *Cotton acreage per capita.* The base data are cotton acreage in 1910 by county. (U.S. Bureau of the Census 1915a.)

- *Tobacco acreage per capita*. The base data are tobacco acreage in 1910 by county. (U.S. Bureau of the Census 1915a.)
- *Parental-background controls*. The mother's and father's occupational income scores 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 assign them incomes of zero. These variables are interacted with census year in the regressions.

For the retrospective-cohort analysis, I focus on state of 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 they 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. Part 3 (table 3) contains enumerated deaths of children under one year. I scale this number by the estimated birth rate (part 1: 482) times the female population (part 1: table 2). (U.S. Bureau of the Census 1894.)
- *Recruits for World War I found rejected for military service because of health "defects," 1917–19*. Fraction of total recruits. (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 or croup, typhoid fever, malaria, diarrheal diseases, and pneumonia. Data are expressed as the fraction of total mortality in 1890. (U.S. Bureau of the Census 1894.)
- *Fertility rate, 1890*. The estimated birth rate (from part 1: 182). (U.S. Bureau of the Census 1894.)
- *Log change in school term length, c1902–32*. Average length of school term, in weeks. (U.S. Office of Education 1905–32.)
- *Log change in average monthly salaries for teachers, c1902–32*. (U.S. Office of Education 1905–32.)
- *Log change in pupil-teacher ratio, c1902–32*. Average attendance divided by number of teachers. (U.S. Office of Education 1905–32.)
- *Log change in school expenditure, c1902–32*. (U.S. Office of Education 1905–32.)

- *Adult literacy rate*. Defined as above.
- *Population urban*. Defined as above.
- *Fraction black*. Defined as above.
- *Male unemployment rate, 1930*. (ICPSR 1984.)

References

- Abbott, Samuel W. 1900. "The Past and Present Condition of Public Hygiene and State Medicine in the United States." Monographs of American Social Economics 19. Boston: Wright and Potter.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369–1401.
- Alabama, Department of Education. Various years (1905–25). *Annual Report*. Montgomery, AL: State of Alabama.
- Arkansas, Superintendent of Public Instruction. Various years (1905–25). *Biennial Report*. Little Rock, AR: State of Arkansas.
- Bleakley, Hoyt. 2002a. "Malaria and Human Capital: Evidence from the American South." Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA, May.
- . 2002b. *Three Empirical Essays on Investment in Physical and Human Capital*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- . 2003. "Disease and Development: Evidence from the American South." *Journal of the European Economic Association* 1 (2-3): 376–86.
- Bleakley, Hoyt, and Fabian Lange. 2004. "Chronic Disease Burden and the Interaction of Education, Fertility, and Growth." Unpublished manuscript, Yale University, New Haven, CT.
- Bogue, Donald J. 1951. *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.
- Bowen, Albert S. 1928. "Activities Concerning Mobilization Camps and Ports of Embarkation. Vol. IV." Prepared under the direction of the Surgeon General, M. W. Ireland. Washington, DC: U.S. Government Printing Office.
- Brinkley, Garland L. 1994. *The Economic Impact of Disease in the American South, 1860–1940*. Ph.D. dissertation, University of California, Davis.
- Dickson, Rumona, Shally Awasthi, Paula Williamson, Colin Demellweek, and Paul Garner. 2000. "Effects of Treatment for Intestinal Helminth Infection on Growth and Cognitive Performance in Children: Systematic Review of Randomised Trials." *British Medical Journal* 320 (June 24): 1697–701.
- Duncan, Otis D. 1961. "A Socioeconomic Index for All Occupations." In *Occupations and Social Status*, ed. A. J. Reiss, 109–38. New York: Free Press.
- Ettling, John. 1981. *The Germ of Laziness: Rockefeller Philanthropy and Public Health in the New South*. Cambridge, MA: Harvard University Press.

- Farmer, Henry F. Jr. 1970. *The Hookworm Eradication Program in the South, 1909–1925*. Ph.D. dissertation, University of Georgia.
- Ferrell, John A., Wilson G. Smillie, Platt W. Covington, and Pauline A. Mead. 1932. *Health Departments of States and Provinces of the United States and Canada*. Public Health Bulletin 184. Washington, DC: Government Printing Office.
- Georgia, Department of Education. Various years (1905–25). *Annual Report*. Atlanta, GA: State of Georgia.
- Havens, Leon C., and Ruth Castles. 1930. “The Evaluation of the Hookworm Problem of Alabama by Counties.” *Journal of Preventive Medicine* 4: 109–14.
- ICPSR (Inter-university Consortium for Political and Social Research). 1984. *Historical, Demographic, Economic, and Social Data: the United States, 1790–1970*. Ann Arbor, MI: ICPSR. Computer file. <http://www.icpsr.org/>.
- Jacocks, W. P. 1924. “Hookworm Infection Rates in Eleven Southern States as Revealed by Resurveys in 1920–1923.” *Journal of the American Medical Association* 82 (20): 1601–02.
- Johnson, Charles S. 1941. *Statistical Atlas of Southern Counties: Listing and Analysis of Socio-Economic Indices of 1101 Southern Counties*. Chapel Hill, NC: University of North Carolina Press.
- Kentucky, Superintendent of Public Instruction. Various years (1905–25). *Biennial Report*. Frankfort, KY: Commonwealth of Kentucky.
- Kofoed, Charles A., and John P. Tucker. 1921. “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* 1 (1): 79–117.
- Kremer, Michael, and Edward Miguel. 2004. “The Illusion of Sustainability.” Working Paper 10324, National Bureau of Economic Research, Cambridge, MA.
- Lebergott, Stanley. 1964. *Manpower in Economic Growth: The American Record since 1800*. New York: McGraw-Hill.
- Louisiana, Superintendent of Public Education. Various years (1905–25). *Biennial Report*. Baton Rouge, LA: State of Louisiana.
- Love, Albert C., and Charles B. Davenport. 1920. *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*. Prepared under the direction of the Surgeon General, M. W. Ireland. Washington, DC: U.S. Government Printing Office.
- Maxcy, Kenneth F. 1923. “The Distribution of Malaria in the United States as Indicated by Mortality Reports.” *Public Health Reports* 308: 1125–38.
- Miguel, Edward, and Michael Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica* 72 (1): 159–217.
- Mississippi, Superintendent of Public Education. Various years (1905–25). *Biennial Report*. Jackson, MS: State of Mississippi.
- NAPP (North Atlantic Population Project). 2004. NAPP: Complete Count Microdata, Preliminary Version 0.2. Minneapolis, MN: Minnesota Population Center. Computer file. <http://www.nappdata.org/>.

- North Carolina, Superintendent of Public Instruction. Various years (1905–25). *Biennial Report*. Raleigh, NC: State of North Carolina.
- Philipson, Tomas J. 2000. “External Treatment Effects and Program Implementation Bias.” Working Paper T0250, National Bureau of Economic Research, Cambridge, MA.
- RSC (Rockefeller Sanitary Commission). Various years (1910–15). *Annual Report*. New York: Rockefeller Foundation.
- Ruggles, Steven, and Matthew Sobek. 1997. Integrated Public Use Microdata Series: Version 2.0. Minneapolis, MN: University of Minnesota, Historical Census Projects. <http://www.ipums.umn.edu/>.
- Sachs, Jeffrey D. 2001. “Tropical Underdevelopment.” Working Paper 8119, National Bureau of Economic Research, Cambridge, MA.
- Smillie, Wilson G., and Donald L. Augustine. 1925. “Intensity of Hookworm Infection in Alabama: Its Relationship to Residence, Occupation, Age, Sex, and Race.” *Journal of the American Medical Association* 85: 1958–63.
- South Carolina, Superintendent of Education. Various years (1905–25). *Annual Report*. Columbia, SC: State of South Carolina.
- Tennessee, Superintendent of Public Instruction. Various years (1905–25). *Biennial Report*. Nashville, TN: State of Tennessee.
- Texas, Superintendent of Public Instruction. Various years (1905–25). *Biennial Report*. Austin, TX: State of Texas.
- U.S. Bureau of the Census. 1894. “Vital and Social Statistics.” In *Compendium of the Eleventh Census, 1890*. Vol. II. Washington, DC: U.S. Government Printing Office.
- . 1915a. “Report on the Statistics of Agriculture in the United States.” In *Thirteenth Decennial Census of the United States 1910*. Vols. 6–7. Washington, DC: U.S. Government Printing Office.
- . 1915b. *Wealth, Debt, and Taxation*. Vol. 2. Washington, DC: U.S. Government Printing Office.
- . 1935. *Financial Statistics of State and Local Governments: 1932*. Washington, DC: U.S. Government Printing Office.
- U.S. Office [Bureau] of Education. Various years (1894–1913). *Annual Report of the Commissioner Education*. Washington, DC: U.S. Government Printing Office.
- Virginia, Superintendent of Public Instruction. Various years (1905–25). *Annual Report*. Richmond, VA: Superintendent of Public Printing.