The Very Long-Run Effect of Large-Scale Deworming in China [*]

Chao Liu[†] Gordon G. Liu[‡]

January 12, 2023

Abstract

This paper studies the long-term impacts of an unprecedented large-scale schistosomiasis control campaign in China in the late 1950s. We present one of the first direct evidence of how an early-life deworming intervention affected life trajectories. The deworming program had a positive effect on the educational attainment of rural people. The effect was larger for people from a low socioeconomic background, suggesting that the program reduced educational inequality. Early-life exposure to the program had substantial effects on labor market success and economic status 50 years later, as well as on the second generation's schooling.

---

[*] We thank Anna Aizer, Richard Freeman, Leander Heldring, Wei Huang, Dean Karlan, Lixing Li, Ameet Morjaria, Matthew Notowidigdo, Nancy Qian, Paola Sapienza, Chuanchuan Zhang, Dandan Zhang, and various seminar participants for helpful comments and suggestions. We thank Xiaoyun Peng for her excellent research assistant work. This paper was previously circulated under the name "Early-Life Health and Lifetime Outcomes: Evidence from the Large-Scale Schistosomiasis Eradication in China."

[†] Liu: Kellogg School of Management, Northwestern University, 2211 Campus Dr, Evanston, IL 60208. Email: chao.liu1@kellogg.northwestern.edu.

[‡] Liu: National School of Development, Peking University, 5 Yiheyuan Rd, Beijing, 100871, China. Email: gordonliu@nsd.pku.edu.cn.

The relationship between early-life health and lifetime outcomes has attracted considerable attention across multiple disciplines, with important implications for public policy.[1] Recent advances in micro-level empirical evidence have been driven by the use of randomized controlled trials and the increased availability of comprehensive administrative data (Baird et al., 2016; Bütikofer and Salvanes, 2020; Bhalotra et al., 2022). Because of poor sanitary and medical conditions in less developed regions, people who need public health interventions the most are precisely the ones who may benefit most from them. However, there is a relatively scarce literature exploring the long-term effects of large-scale public health interventions in developing countries, likely due to a lack of data on the interventions and treated individuals.[2] Therefore, examining the impacts of the improvement in early-life health at different points of the life course in low-income areas remains important for research and policy concerns. This paper contributes to filling this gap by studying the long-term effects of a large-scale deworming campaign in China targeted at schistosomiasis in the late 1950s.

Schistosomiasis is a parasitic disease that is widespread in more than 70 nations and territories, posing a significant threat to the socioeconomic development of tropical and subtropical regions. Naghavi et al. (2017) reveals that its global prevalence and burden of illness are even greater than those of malaria. Due to its chronic and insidious properties, schistosomiasis has a long-term effect on children's development. Maternal schistosomiasis results in low birth weight and preterm delivery, which can impair infants' subsequent cognitive and physical development. Early-life infections can directly cause poor growth and learning difficulties (Ezeamama et al., 2018).

China was the largest endemic area of this disease in the first half of the 20th century. In 1956, Mao Zedong issued the famous slogan "Must Eliminate Schistosomiasis." Following Mao's strong determination, the eradication program started nationwide and mainly implemented snail control to reduce parasite transmission. This deworming campaign has three unique features that facilitate our empirical analysis. First, the program achieved great success sufficiently long ago, offering us a unique opportunity to track the lifetime effects of a massive and persistent reduction in schistosomiasis prevalence. Second, the deworming campaign was mobilized in a top-down manner, so the public health intervention was less likely to be related to the expected economic growth in endemic areas. Finally, the treatment was not particularly designed for a specific group of people, so we can explore the critical period

---

[1] See Almond, Currie and Duque (2018) for a recent and thorough review.

[2] Existing research in this strand mainly examines the effects of malaria eradication campaigns (Barreca, 2010; Cutler et al., 2010; Lucas, 2010; Barofsky, Anekwe and Chase, 2015).

of deworming programs.

To conduct the analysis, we hand-collect pre-control schistosomiasis infection rates at the county level from a Chinese disease atlas. We combine the disease data with two individual-level datasets. The first one is the 1990 China Population Census, which allows us to estimate how the deworming campaign affected rural people's educational attainment and employment status in their thirties. Although we do not observe people's migration histories in the census, this barely affects our estimation as migration was highly restricted in China at that time due to the household registration (*hukou*) system. The second individual-level dataset is the 2010 China Family Panel Studies (CFPS), which offers three advantages over the census for our analysis. First, we can construct a refined sample for analysis by directly controlling for migration and family background.[3] Second, it allows us to examine the impact of the public health intervention that occurred over 50 years ago on a variety of adult outcomes. Third, we can link individual-level information across three generations to explore potential intergenerational effects. Our main identification strategy is a difference-in-differences (DID) model with county and cohort fixed effects, which compares lifetime outcomes of people born before and after the deworming program across areas with different pre-control schistosomiasis infection rates.

We conduct a comprehensive inquiry into the long-run effects of childhood exposure to the schistosomiasis control campaign on multiple lifetime outcomes at different stages of the life course. We show that the deworming campaign had a significantly positive effect on educational attainment. According to our baseline estimates from the census sample, rural men received roughly 0.11 more years of schooling from a drop of 10 percentage points in the schistosomiasis infection rate, and rural women gained about 0.19 more years of schooling. Estimates using the CFPS data confirm these results. Although the campaign had no effect on the employment status of people in their thirties, it greatly increased the possibility of having a job for females in their fifties. Our estimates using the CFPS sample also suggest that the disease control program improved adult economic status over 50 years after the treatment. For rural males, reducing the schistosomiasis infection rate from 10 percent to zero would lead to a 10.59 percent increase in per capita household income, a 10.44 percent increase in per capita household consumption, and a 13.82 percent increase in per capita net worth. Unlike the education effect, the economic effect was smaller

---

[3] The CFPS data were collected in 2010. Internal migration was not rare at that time, so we directly address the migration concern when we use this sample.

for treated females. In addition to educational improvement, we show that labor market success (having a more prestigious job for males and being employed for females) and adult health condition (only for males) are potential channels between early-life health and adult income. Using the CFPS dataset, we find that the positive effect on educational attainment was larger for people from a low socioeconomic background, suggesting that the deworming campaign not only raised the overall level of rural education but also reduced socioeconomic inequality. The positive effect of this public health intervention on rural education lasted more than one generation. If a woman experienced a reduction in the schistosomiasis infection rate of 10 percentage points during her childhood, her first child was 8.64 percentage points more likely to complete high school.

We address a wide range of threats to identification. First, we conduct a cohort analysis to show that there are no differential pre-trends in outcomes of interest with respect to the schistosomiasis infection rate. Second, we directly control for local intensity of the Great Famine, the Send-Down Movement, and the Cultural Revolution to rule out alternative explanations of these concurrently influential events. Third, to address the concern about the measurement error of treatment intensity, we use the proportion of water area in a county as the instrument variable (IV) for the pre-control schistosomiasis infection rate. The first stage relies on the fact that the spread of schistosomiasis depended on infected water in most endemic areas in China. Furthermore, we provide a test to show the plausibility of this IV strategy. The idea is as follows: if schistosomiasis was the only channel through which the ratio of water area affected the outcomes of interest, the relationship between the dependent variable and the instrument variable should not exist in uninfected areas. A significant reduced-form IV result in infected areas and an insignificant result in uninfected areas together boost our confidence in this IV strategy.[4] Finally, we argue that our results at least provide a lower bound of the true treatment effect, even if selective mortality exists. Our results are also robust to a suite of robustness checks. These include specification tests that make the parallel trends assumption more likely to be satisfied, an alternative DID estimator that allows heterogeneous treatment effects (De Chaisemartin and d'Haultfoeuille, 2020), placebo tests in the sample of unaffected people, and randomization inference (Bertrand, Duflo and Mullainathan,

---

[4] Previous papers use other geographical factors, such as temperature and precipitation, as instrument variables for disease prevalence (Cutler et al., 2010; Lucas, 2010). Our test is also useful in those settings because the disease was only prevalent in some areas, but the instrument variable existed in all areas.

2004).

This paper provides new empirical evidence for understanding the gender heterogeneity in the long-term effects of early-life health improvement. Pitt, Rosenzweig and Hassan (2012) develop a model of human capital investment and activity choice, which predicts that public health interventions will increase more years of schooling for women relative to men in a brawn-based economy. The Chinese economy in our research background is similar to the model setup, where brawn was the major production input as most people in our sample were farmers. Our estimates from both the census and the CFPS sample suggest that the education effect for women was greater than that for men. Moreover, we provide empirical evidence consistent with the implications of the economic theory: shorter men will have more education than taller men, but taller women will have more education than shorter women after the public health intervention. Our study also sheds new light on the critical period of deworming programs. While considerable efforts have been made to reduce infections and morbidities in sub-Saharan Africa, these efforts seem to be tailored only towards school-aged children. At the time of treatment, program participants have already passed the age window considered most critical for early childhood development.[5] Our evidence suggests that fetal interventions to reduce schistosomiasis infection may be more effective than interventions during pre-school or school age, which is consistent with the literature on neurological development.

This paper contributes to the ongoing debate on the effectiveness of mass deworming programs. Existing papers suggest that school-based deworming interventions in Africa are cost-effective for improving school participation and labor market outcomes (Miguel and Kremer, 2004; Baird et al., 2016; Ozier, 2018).[6] However, a medical review argues that there is substantial evidence that mass deworming programs do not improve average nutritional status, hemoglobin, cognition, school performance, or survival (Taylor-Robinson et al., 2015). In this paper, we complement the literature by evaluating the long-term effects of a society-wide schistosomiasis eradication program in China.[7] The scale of the intervention and the number of potentially treated

---

[5] Nowakowski (2006) finds that the brain doubles in size in the first year. By age three, a child's brain has reached 80% of its adult volume.

[6] These papers cite Croke (2014) as supportive evidence on the effectiveness of deworming programs. The original working paper that uses data from 2010 and 2011 survey rounds finds higher mathematics and English scores for children given deworming medication early in life in Uganda (Croke, 2014). However, the published version uses additional four survey rounds (2012-2015) and finds that the deworming program resulted in no statistically significant gains in numeracy or literacy 7-12 years after the treatment (Croke and Atun, 2019).

[7] In terms of pathology, schistosomiasis is caused by water-transmitted worms. Previous deworm-

people were unprecedented. Because the treatment status was determined by location and cohort, our estimates naturally incorporate direct effects and externalities of the intervention. Our study is most related to Bleakley (2007), which finds that the eradication of hookworm in the southern United States had a positive effect on school enrollment, literacy, and adult income.[8] To the best of our knowledge, we are among the first to study the long-run effects of in utero deworming programs. Together with other recent research on deworming, this paper helps to paint a complete picture of the long-run benefits of deworming programs in developing countries.

In addition to our specific contribution to the deworming literature, this paper fits into a large and expanding literature on the fetal origins hypothesis. The recent literature exhibits an increasing interest in examining the long-run effects of in utero interventions on adult outcomes. Latest examples include evaluating the impact of extra medical care at birth in Chile and Norway (Bharadwaj, Løken and Neilson, 2013), the phaseout of leaded gasoline in Sweden (Grönqvist, Nilsson and Robling, 2020), and an infant health intervention in Sweden (Bhalotra et al., 2022). Policymakers, especially in low-income regions, hope to know whether certain intervention tools derived from scientific laboratory experiments or randomized controlled trials can generate their anticipated results when scaled up and implemented through government policies.[9] However, previous literature on this topic mainly studies the negative shocks to early-life health in developing countries.[10] We contribute to this literature by investigating the long-term effects of a large-scale public health intervention in an extremely underdeveloped economy.[11] Even though there were inadequate medical facilities and a turbulent political environment in China at that time, we document lasting and substantial effects of the program, so our results should be encouraging for currently endemic areas. Currie and Vogl (2013) highlight that the literature still has relatively limited understanding of heterogeneity in the long-run effects, the critical periods of different interventions, and possible intergenerational effects. As discussed above, our paper provides new empirical evidence for these unanswered questions.

The remainder of this paper proceeds as follows. Section I introduces the disease and the deworming program in China. Section II outlines the data and empirical

---

ing literature focuses on diseases caused by soil-transmitted worms.

[8] The treatment in eleven southern states were also targeted at school-aged children.

[9] See Banerjee et al. (2017) for a detailed discussion.

[10] For example, related Chinese studies mainly focus on the effects of the 1959-1961 Great Famine (Chen and Zhou, 2007; Almond et al., 2007; Meng and Qian, 2009).

[11] By 1950, per capita GDP in China was merely 21 percent of the world average (Dollar, Huang and Yao, 2020).

strategies employed in this paper. Sections III and IV present our empirical results from the census and the CFPS data, respectively. Section V concludes.

I Background

A Schistosomiasis and Its Spread in China

Schistosomiasis is an acute and chronic parasitic disease caused by blood flukes (trematode worms) of the genus Schistosoma. People get infected when larval forms of the parasite penetrate the skin through infested water. Lack of clean water and poor sanitation make schistosomiasis particularly prevalent in rural communities in tropical and subtropical areas. Typical acute symptoms include abdominal pain, diarrhea, bloody stool, and blood in the urine. Without prompt and adequate medical treatment, cases may present as advanced schistosomiasis with symptoms such as splenomegaly, ascites, colonic tumoroid proliferations, or growth retardation (Colley et al., 2014). Long-time adult patients may develop kidney failure, liver damage, and even bladder cancer. Child patients may suffer from poor growth and learning difficulties (World Health Organization, 2017). Patients with schistosomiasis receive insufficient treatment since it is difficult to identify this disease in its early stage. At least 258 million people were estimated to need treatment in 2014, but only 61.6 million were reported to receive treatment (World Health Organization, 2016).

Archeological evidence has demonstrated the long history of schistosomiasis in China. An ancient Chinese medicine book (circa 400 BC) recorded similar clinical symptoms of acute schistosomiasis. Archeologists detected schistosome eggs in a female corpse that can be traced back to circa 100 BC (Mao and Shao, 1982). In the early 20th century, China was the largest endemic area of *S. japonicum* infections.[12] The disease was spread across 12 province-level regions in the south of China and along the Yangtze River: Anhui, Fujian, Guangdong, Guangxi, Hunan, Hubei, Jiangsu, Jiangxi, Shanghai, Sichuan, Yunnan, and Zhejiang. There were three types of endemic areas based on terrain: marshland and lake regions, hilly and mountainous regions, and plain regions with waterway networks. In the early 1950s, the estimated number of patients was over 10 million (Mao and Shao, 1982). The prevalence of

---

[12] The three main species infecting humans are *S. haematobium*, *S. japonicum*, and *S. mansoni*. Three other species, more localized geographically, are *S. mekongi*, *S. intercalatum*, and *S. guineensis* (previously considered synonymous with *S. intercalatum*). *S. haematobium* occurs in Africa and the Middle East. *S. mansoni* occurs in Africa, the Middle East, the Caribbean, Brazil, Venezuela and Suriname. *S. japonicum* only occurs in China, Indonesia and the Philippines.

schistosomiasis greatly impeded economic development and improvement of living standards. However, efforts to control the disease were limited at that time due to poverty, internal strife, and lack of effective tools (Chen et al., 2016).

B Deworming Program in the 1950s

Chinese scientists started to carry out epidemiological surveys to determine the prevalence of *S. japonicum* infections in the early 1950s. In 1956, Mao Zedong called for all party members and Chinese people to eradicate schistosomiasis. The central government then established a bureau to lead the national schistosomiasis control and issued an instruction on the eradication program. A deworming campaign was carried out in the 12 endemic provincial administrative units shortly thereafter. The launch of this national campaign heavily relied on the supreme leader's willingness, and it was mobilized in a top-down manner, so it provides an exogenous shock to schistosomiasis infections. Special anti-schistosomiasis health stations were established at three levels in endemic areas (province, county, and community). Figure 1 shows the number of newly established medical institutions that specialized in schistosomiasis from 1950 to 1985. We see a spike in 1956, which was the start of the schistosomiasis control campaign. Health workers in medical schools and cities were dispatched to help in rural endemic areas, but most of the eradication effort was extended through a "barefoot doctor" system.[13] The governments in endemic areas usually bore the cost of treatment, so rural population could receive free treatment. The key tool to eliminate schistosomiasis was snail control through environmental modification and mollusciciding (Li et al., 2016). Snail control can interrupt the life cycle of Schistosoma in its snail host to reduce parasite transmission. Chemical repellents, safe water, and sanitary toilets were provided as complementary interventions. A hygiene movement and health education were also used as preventive measures (Xu et al., 2016).

[Insert Figure 1 here]

The eradication efforts resulted in a great decline in schistosomiasis infections immediately. Thirteen formerly endemic counties had eradicated the disease by the end of 1957. In Hubei province, over 140,000 patients received treatment and more

---

[13] Barefoot doctors were special medical workers in rural areas. They received minimal basic medical and paramedical training. In addition to treating common diseases, they provided basic preventive health care and family planning.

than 80 percent of them recovered within several years after the launch of the disease control program. By 1989, about 72 percent of the formerly endemic counties had succeeded in interrupting and controlling the disease transmission, which proves the effectiveness of the schistosomiasis control campaign in China (Yuan, Zhuo and Zhang, 1989).

II Data and Empirical Strategy

A Data

We compile our data from three main sources. To measure schistosomiasis prevalence, we collect data from the China Schistosomiasis Atlas (Qian, 1988). This book records detailed information on the number of patients, infection rates, and snail habitat areas at the county level before and after the deworming campaign.[14] We use the ratio of the number of patients to total population as our measure for pre-control schistosomiasis prevalence. Our final treatment sample contains 255 endemic counties in 12 endemic provincial administrative units. Uninfected counties in these provinces make up our control group. Figure 2 shows the regional variation in pre-control disease prevalence and the success of the deworming campaign.

[Insert Figure 2 here]

We then match the disease data with two individual-level datasets. The first one is the 1 percent sample of the 1990 China Population Census. This dataset contains basic demographic variables we use in this paper: gender, year of birth, ethnicity, *hukou* type (urban or rural residential registration), county of *hukou*, education, and employment status for respondents over 16. We generate a new variable that measures years of schooling based on the original categorical education variable and the status of completion.[15] One potential shortcoming of the census data is the imprecise age information, as people tend to round their age to numbers ending in 0 or 5.[16] We directly test whether this problem exists in the 1990 China Population Census. Figure

---

[14] Pre-control data were collected in the early 1950s, and post-control data were collected in 1981. Section B of the Appendix introduces this disease dataset in detail.

[15] Specifically, the mapping is: illiterate = 0 years, (in)complete primary school = 6 (3) years, (in)complete junior high school = 9 (7.5) years, (in)complete senior high school or specialized secondary school = 12 (10.5) years, and (in)complete college or junior college = 16 (14) years.

[16] The age-heaping problem is common in developing countries' census data. Cutler et al. (2010) and Barofsky, Anekwe and Chase (2015) group people into five-year birth cohorts to overcome this problem.

A1 plots the distribution of the last digit of the reported age in the rural population. The distribution is fairly even, which is consistent with previous research on the accuracy of age data in the China Population Census (Huang, 2009). Therefore, we stick to the reported birth year in our analysis. Following Chen et al. (2020), we assume the county of residence in 1990 as the county of birth and the *hukou* type in 1990 as the childhood living area type. This is not a strong assumption in our context. Due to the household registration (*hukou*) system, migration between counties was scant in rural areas, and there were rare cases of people changing their *hukou* type at that time.[17] As schistosomiasis was mainly endemic in rural areas, we limit our sample to people with rural *hukou* and born between 1946 and 1966 (normally these people had finished their education in 1990). Table 1 provides the summary statistics of the census sample. The average age of the people in this sample was about 33. Men had a substantially higher education level than women. About 92 percent of the sample were of Han ethnicity. Because most people were in their prime age and the great majority of them were agricultural workers, the employment rate was extremely high.

[Insert Table 1 here]

The second individual dataset is the 2010 wave of the China Family Panel Studies (CFPS). We use this dataset to replicate our results from the census sample, estimate the deworming effect in a longer term, and explore intergenerational impacts. Although the CFPS does not contain people's birthplaces, it provides useful information so that we can directly address the migration concern. Specifically, we drop individuals who moved out of their birthplaces and keep people holding rural *hukou* when they were 3 years old.[18] We keep respondents born between 1940 and 1970 in endemic provinces plus 4 adjacent uninfected provinces, which leaves us with a sample including 23 infected and 89 uninfected counties in 17 provinces. In Section IV.A, we show that the resulting sample remains representative. Summary statistics of key variables are reported in Table 2. Compared with the census sample, one noticeable change is that the average age rose to 52. Roughly 64 percent of men and 51 percent of women were employed as the elder cohorts had retired in 2010. The average

---

[17] The *hukou* system was introduced in 1958 as a response to the large influx of people from the countryside to the cities during the Great Leap Forward. Migration was effectively stopped until the early 1990s.

[18] We use a restrictive version of the CFPS data that has information on individual's current residence county, but even this version does not provide individual's birth county. Thus, we decide to drop movers and argue that this will not bias our analysis.

educational level was slightly lower than that in the census sample. The average per capita household income was approximately 7,700 CNY ($1,137), and about 73% of the income was used for consumption. The average per capita net worth was about 52,000 CNY ($7,680), which mainly reflected the value of land asset.

[Insert Table 2 here]

B Empirical Strategy

This paper follows the identification strategy used in Bleakley (2010), Cutler et al. (2010), and Lucas (2010) to identify the long-term effects of the schistosomiasis control program. We run regressions using the following form:

$$y_{ijc} = \alpha + \beta \left(sch_j \times post_{ic} \right) + X_{ijc}\Theta + \delta_j + \delta_c + \varepsilon_{ijc} \tag{1}$$

where y_{ijc} is an adult outcome of individual i in county j, who belongs to cohort c. sch_j is the pre-control schistosomiasis infection rate in county j, measuring the treatment intensity. $post_{ic}$ is a dummy variable indicating that individual i was born after 1957. X_{ijc} is a set of control variables, which could include ethnicity and parental education levels. δ_j and δ_c are county fixed effects and cohort fixed effects, respectively. The coefficient of interest is β, which captures the effect of the schistosomiasis control program on lifetime outcomes. To explore potential heterogeneous effects by gender, we estimate equation (1) for rural males and females separately.

As discussed in Bleakley (2007), this empirical strategy assumes that (i) there were regional variations in schistosomiasis exposure, and (ii) areas with higher infection rates had higher treatment intensity. Panel A of Figure 2 validates the first assumption, showing large variations in pre-control schistosomiasis prevalence at the county level. Figure 3 illustrates that areas with higher pre-control infection rates indeed saw a greater decline in schistosomiasis prevalence, which supports the second assumption. If a county (represented by a dot in the figure) eradicated schistosomiasis, it will be located on the diagonal line. As shown in Figure 3, most endemic counties are on or close to the diagonal line.

[Insert Figure 3 here]

Our empirical strategy implicitly assumes that there are no differential changes in outcomes between different counties with varying infection rates in the absence of the schistosomiasis control program. In this part, we provide some suggestive evidence on the validity of this parallel trends assumption.[19] Figure A2 plots the cohort-level mean residual outcomes conditional on ethnicity and county fixed effects for rural women (panel A) and men (panel B) born between 1946 and 1966. Different colored lines denote the average residual outcomes of counties in four quartiles of the pre-control infection rate. The black line denotes the average residual outcomes in uninfected counties. We focus on educational attainment and employment status in the census sample. To measure educational attainment, we use years of schooling and three dummy variables indicating literacy, primary school completion, and junior high school completion. As shown in this figure, the treatment and control groups move in a parallel fashion before 1957, but they diverge (except the literacy rate and employment rate) after the deworming program.

III Empirical Results from the Census Data

A Baseline Results

We first investigate the effects of the deworming campaign on educational attainment. Table 3 reports our baseline results by estimating equation (1). Panel A shows that if a county with a 30 percent pre-control infection rate eradicated schistosomiasis, rural women in this county would be 2.97, 6.84, and 6.06 percentage points more likely to be literate, complete primary school, and complete junior high school, respectively.[20] On average, rural women in this county would have an additional 0.56 years of schooling. We find that rural men received smaller educational benefits from the reduction in schistosomiasis infections. Panel B of this table demonstrates that the same decrease in the pre-control infection rate led to an increase of 0.34 years in years of schooling and an increase of 3.60 (5.10) percentage points in the probability of completing primary (junior high) school among rural men. However, the disease control program did not improve male literacy.[21] One explanation for this result is that there was less scope to increase this outcome, as about 89 percent of rural men born before the deworming campaign could read and write.[22] It is worth noting that our baseline

---

[19] More formal analysis will be presented in section III.

[20] For reference, the 90th percentile of the pre-control schistosomiasis infection rate among endemic counties was 30.2 percent.

[21] Selective mortality may also explain this finding. See section III.B for the discussion.

[22] Similarly, Bleakley (2007) finds that the hookworm eradication in the American South had little impact on school enrollment and literacy for white people. He argues that this is because whites had less scope to increase these measures of human capital investment.

estimation is very likely to be a lower-bound of the true treatment effect, since we assume that rural people did not relocate. Although this is not a strong assumption given the rigid *hukou* system, it is true that the best-educated rural people were more likely to migrate and switch *hukou* type (Zhao, 1997). Therefore, our sample may not include those best-educated rural children, which makes our model underestimate the positive effect of the disease control program on educational attainment.

Using the census data, we also examine the effect of the deworming campaign on employment status when people were in their thirties. For both men and women in rural areas, we are unable to reject the null hypothesis that the changes in employment rate across counties with different schistosomiasis infection rates were the same. This is consistent with the raw comparison in Figure A2, as we do not detect a divergence in employment between treatment and control groups after 1957. Since most patients were not deprived of working ability completely by schistosomiasis, they were still counted as farmers in the census data. As shown in Table 1, more than 99 percent of men and 93 percent of women were employed, which left little room for further increase in this outcome.

[Insert Table 3 here]

To present the estimated effect graphically, we conduct a cohort analysis using the following specification:

$$y_{ijc} = \alpha + \sum_{c} \beta_c \left(sch_j \times cohort_{ic} \right) + X_{ijc}\Theta + \delta_j + \delta_c + \varepsilon_{ijc} \tag{2}$$

where $cohort_{ic}$ equals 1 if individual i belongs to cohort c. Equation (2) allows the effects of the schistosomiasis control program to vary by cohorts. We expect that the coefficients associated with the pre-control cohorts are small and insignificant, but those associated with the post-control cohorts are economically and statistically significant. The effect on the cohort born in 1956 is normalized to zero, so all other coefficients represent the treatment effects relative to people born in 1956. Panel A of Figure 4 plots the cohort-specific coefficients β_c for rural females. The coefficients fluctuate around zero for cohorts born before 1957, which supports the assumption of parallel trends across counties with different pre-control schistosomiasis prevalence. Except when we use literacy or employment status as the dependent variable, most coefficients after 1957 are large in magnitude and statistically significant. This confirms our baseline results in Table 3 that the deworming campaign had a positive impact on female educational attainment. The break in the coefficient series, which is coincident with the timing of the schistosomiasis control campaign, demonstrates that it is reasonable to treat people born before 1957 as pre-control cohorts. Panel B of this figure shows the same cohort analysis for rural males. We find robust positive effects on years of schooling and the likelihood of completing junior high school. Similar to the baseline DID estimation, the education effect on males was smaller than that on females.

[Insert Figure 4 here]

B Threats to Identification

In this part, we discuss several threats to our baseline analysis and how we address these concerns. We show that these threats will not overturn the above findings and our baseline results underestimate the true treatment effects in most cases.

Concurrent Events.— During our study period, China experienced several influential events, such as the Great Famine, the Great Leap Forward, and the Cultural Revolution. Admittedly, these events had direct impact on educational attainment and employment status. Moreover, there might be a linkage between local disease burden and the severity of these historical events. To address this concern, we directly control for the possible confounding effects of contemporaneous events in a flexible way.

We first introduce how we measure these concurrent events intensity at the county level.[23] Following Meng and Qian (2009), we define local famine severity as one minus the ratio of the cohort size of the famine cohorts (1959-1961) to that of the non-famine cohorts (1955-1957). To measure the intensity of the Cultural Revolution, we use the ratio of victims to the county population in 1964.[24] During the Cultural Revolution, the Send-Down Movement mandated about 16 million urban youth to temporarily resettle in rural areas. This event seems to be a particular threat to our identification, as it happened in rural areas at almost the same time. Following Chen et al. (2020),

---

[23] Due to data limitation, we measure the intensity of the Great Leap Forward at the province level. Following Meng, Qian and Yared (2015), we use the increase in steel production, participation rate in communal dining halls, magnitude of the Anti-Rightest purges, and the first principal component of these three variables as proxies for political zealousness at the province level. Table A1 shows that none of these measures is correlated with pre-control schistosomiasis infection rate at the province level.

[24] The data come from the China Political Events Dataset (Walder, 2014).

we use the ratio of received send-down youths to the county population in 1964 to measure its intensity.

Formally, we estimate the following equation:

$$y_{ijc} = \alpha + \beta\left(sch_j \times post_{ic}\right) + \sum_c \gamma_c \left(GF_j \times cohort_{ic}\right) + \sum_c \tau_c \left(CR_j \times cohort_{ic}\right) + \sum_c \lambda_c \left(SD_j \times cohort_{ic}\right) + X_{ijc}\Theta + \delta_j + \delta_c + \varepsilon_{ijc} \quad (3)$$

where GF_j, CR_j, and SD_j measure the severity of the Great Famine, the Cultural Revolution, and the Send-Down Movement in county j as we described above. Equation (3) allows the effects of three influential events to vary by cohorts. Table 4 reports that the estimated coefficients are similar to our baseline estimates in magnitude and statistical significance. These results show that our baseline estimates do not suffer from the confounding effects of other concurrent historical events.

[Insert Table 4 here]

Furthermore, we rule out alternative stories of other events by exploring potential mechanisms behind the increase in educational attainment. As discussed in section I, the deworming campaign only applied various health and medical tools to control the disease. If our findings were due to the deworming campaign, health improvement should be the primary channel. In contrast, other historical events (e.g., the Great Leap Forward and the Send-Down Movement) were more likely to have an impact through educational infrastructure. We investigate the effect of the deworming campaign on the expansion of educational infrastructure.[25] Table A2 shows that the number of primary schools or secondary schools did not increase in endemic counties after the schistosomiasis control program. Therefore, our findings cannot be fully explained by the Great Leap Forward or the Send-Down Movement.

Measurement Error.— Because the disease prevalence data were collected in a bottom-up way, there may be some concern about whether the pre-control infection rate truly reflects the geographic variations in pre-control schistosomiasis prevalence. On one hand, low-endemic counties might have incentives to report a higher infection rate to receive more fiscal support from upper-level governments. On the other hand, high-endemic counties might have incentives to report a lower infection rate

---

[25] We use the number of primary schools and secondary schools as dependent variables and estimate a DID model at the county level.

to hide the large disease burden. Our ordinary least squares (OLS) estimates will be downward biased in both cases. Since we already have obtained significantly positive results, the above estimates at least provide a lower bound of the positive effect of the schistosomiasis control program in rural China. To evaluate the treatment effect more precisely, we next exploit an instrument variable strategy (IV). The transmission of schistosomiasis relied on infected water in most endemic counties in China, so we use the county's proportion of water area as an instrument variable for the pre-control schistosomiasis infection rate. The first and second stage of the IV regression are given by equations (4) and (5):

$$D_{ijc} = sch_j \times post_{ic} = \alpha + \rho \left(water_j \times post_{ic} \right) + X_{ijc}\Theta + \delta_j + \delta_c + \varepsilon^1_{ijc} \tag{4}$$

$$y_{ijc} = \alpha + \beta \hat{D}_{ijc} + X_{ijc}\Theta + \delta_j + \delta_c + \varepsilon^2_{ijc} \tag{5}$$

where $water_j$ is the share of water area in county j. All other terms are as previously defined.

Table A3 presents the first stage result, which shows that our instrument variable is highly correlated with the endogenous independent variable. This IV strategy also rests on the assumption that schistosomiasis prevalence is the only channel through which the proportion of water area can affect the outcomes of interest in endemic counties. If this exclusion assumption holds, the relationship between the instrument variable and the dependent variable that exists in infected areas should not exist in uninfected areas. To provide supportive evidence for the validity of our instrument variable, we run reduced-form IV regressions in infected and uninfected areas separately. Table 5 presents the reduced-form estimation results. Panels A and B reveal a significantly positive relationship between the instrument variable and the dependent variables in endemic counties. In contrast, panels C and D report small and insignificant coefficients when we run the same reduced-form IV regressions in four uninfected provinces adjacent to the 12 endemic provinces.[26] This falsification test provides suggestive evidence for the validity of our instrument variable.

[Insert Table 5 here]

After presenting the plausibility of our instrument variable, we then estimate the two-stage least squares model (2SLS) in a sample consisting of marshland, lake regions, and plain regions with waterway networks. The large F statistics reported in

---

[26] Four adjacent but uninfected provinces are Henan, Shandong, Shaanxi, and Gansu.

Table 6 again suggest a strong first stage. As shown in this table, the IV coefficients are indeed larger in magnitude than their corresponding OLS coefficients. Overall, we find very similar results as our OLS estimates. The schistosomiasis control campaign had a positive effect on educational attainment, but it had limited impact on male literacy and people's employment status in their thirties. According to column 1 of this table, rural men gained roughly 1 year more schooling from a reduction of 30 percentage points in the schistosomiasis infection rate, and rural women gained another extra 0.3 years.

[Insert Table 6 here]

In addition, we conduct an IV version of the cohort analysis. Specifically, we use $water_j \times cohort_{ic}$ as the instrument variable for $sch_j \times cohort_{ic}$ in equation (2). Panel A of Figure A3 plots the cohort-specific coefficients for rural women, which has a similar pattern as the OLS version of the cohort analysis. Most coefficients before 1957 are economically and statistically insignificant. In contrast, the coefficients after the deworming campaign are large in magnitude and statistically significant except when the dependent variable is literacy or employment status. For instance, there is a discrete increase in the probability of completing junior high school after 1957, and the magnitude of this increase is similar to the coefficient in Table 6. Panel B of Figure A3 also suggests that the schistosomiasis control campaign had a significantly positive effect on years of education and the likelihood of completing junior high school among rural males.

Selective Mortality.— Selective mortality is another common threat to studies estimating long-run effects of early-life health shocks. It is a process whereby disadvantaged individuals die at younger ages than their more advantaged peers. Therefore, the individuals in the census sample may appear healthier and more educated than they would if the more disadvantaged individuals did not die early. We address this concern in two steps. First, the schistosomiasis mortality rate is only approximately 1 percent of the malaria mortality rate, so selective mortality is a minor problem in our setting (Naghavi et al., 2017). Second, even if selective mortality exists in our research, our results at least provide a lower bound of the positive effect of the schistosomiasis control campaign, as long as the campaign reduced mortality.[27] Because schistosomiasis has a relatively low mortality rate, the deworming program only saved

---

[27] Li and Wei (2017) use aggregate data at the city level to find that the treatment and cure of schistosomiasis reduced mortality in endemic areas.

the least healthy people. Given that health is positively correlated with educational attainment and labor participation, the least healthy people now included in the census sample will pull down the average years of schooling of the post-control cohorts. This may also explain why we find a weak effect on literacy rate. In a world without the deworming campaign, people who are weak in health and intelligence will not exist in the treatment group. However, in a world with the deworming campaign, they will survive and lower the average literacy rate of the treatment group.

C Discussion and Robustness Checks

Previous literature evaluates the short-run and long-run effects of deworming programs targeted at school-aged children (Miguel and Kremer, 2004; Bleakley, 2007; Baird et al., 2016), but little is known about the effects of deworming programs targeted at other strata of the population at risk, particularly infants and preschool children.[28] The deworming campaign in China in the late 1950s provides a unique opportunity to answer this question. Unlike school-based deworming programs, the large-scale deworming campaign in this paper was not tailored towards a specific group. Based on this feature, our empirical framework implicitly tests the fetal origins hypothesis, as we classify people born after the deworming campaign as the treatment group. To the best of our knowledge, we present one of the first direct evidence of the long-run impacts of reducing schistosomiasis infections in the maternal, fetal, and neonatal period.[29] To find the critical period of schistosomiasis treatment, we estimate the following specification:

$$y_{ijc} = \alpha + \beta_1 \left(sch_j \times cohort^1_{ic} \right) + \beta_2 \left(sch_j \times cohort^2_{ic} \right) + X_{ijc}\Theta + \delta_j + \delta_c + \varepsilon_{ijc} \quad (6)$$

where $cohort^1_{ic}$ equals 1 if individual i was born between 1957 and 1966, and $cohort^2_{ic}$ equals 1 if individual i was born between 1952 and 1956. People born between 1942 and 1951 are left as the reference group. Thus, cohort 1 got treatment in utero, cohort 2 got treatment during early childhood (1-5 years old), and the reference group got treatment when they were school-aged children (6-15 years old). Table 7 reports the

---

[28] Recent medical research has used animal models and case reports to detect the link between maternal schistosomiasis and adverse pregnancy outcomes (Ajanga et al., 2006; Godwin et al., 2010; McDonald et al., 2014).

[29] Ozier (2018) finds a large long-term cognitive effect for treated children who were less than one year old. when school-age children in their communities received mass deworming treatment. The author exploits the externalities from a randomized deworming intervention aimed at school-age children.

estimated results of equation (6). We find that the fetal disease environment was extremely crucial for educational attainment. The coefficients of the first interaction term ($sch_j \times cohort^1_{ic}$) are large and consistently significant, indicating that treatment in utero was more effective than treatment in school age.[30] We cannot reject the null hypothesis that the treatment in early childhood and in school age had the same effect. Our findings support the recent recommendation by medical experts to prevent infection before and during pregnancy (Salawu and Odaibo, 2014).

[Insert Table 7 here]

We summarize robustness tests in Table A5. First, we perform two specification tests. In column 1, we add a linear county-specific time trend, so this specification allows different counties to have different growth paths. One concern on our identification strategy is that the increase in educational attainment could be an artifact of mean reversion. To address this concern, we include a series of interaction terms between the average of the dependent variable among people born from 1946 to 1956 and cohort dummies (i.e., $\bar{y}_{j,46-56} \times cohort_{ic}$). As shown in the first two columns, our results are robust to these modifications of the baseline model.

Second, we estimate an alternative DID estimator that allows heterogeneous treatment effects. Recent applied econometrics literature shows that two-way fixed effects estimations of DID coefficients can lead to substantial biases when there are heterogeneous treatment effects, even if the parallel trends assumption is satisfied (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). As the treatment status is continuous in our research, we follow De Chaisemartin and d'Haultfoeuille (2020) to estimate a new DID estimator, DID_m, that is still valid if the treatment effect is heterogeneous. In Appendix C, we derive the estimator and discuss the satisfaction of the identification assumptions in our setting. Column 3 of Table A5 displays our estimates of DID_m. The results are similar to the cohort analysis and IV estimates. We find a significantly positive impact of the deworming campaign on years of schooling and the likelihood of completing primary and junior high school. The positive education effect on males was slightly smaller than that on females.

---

[30] Treatments in early childhood and in school age were also effective. To see this, we add another interaction term $sch_j \times cohort^3_{ic}$ to equation (6), where $cohort^3_{ic}$ equals 1 if individual i was born between 1942 and 1951. People born between 1935 and 1941 are used as the reference group. The effect on the reference group was mechanically zero because these people were over the normal age of completing junior high school when the deworming campaign started. Table A4 shows that treatment was also effective in early childhood and in school age.

Third, we conduct two placebo tests. In column 4, we estimate the baseline model in a sample of people with urban *hukou* in 12 endemic provinces. Because schistosomiasis was mostly prevalent in rural areas where people lacked safe water sources, we expect a small and insignificant effect of the deworming program in the urban sample. In column 5, we assume that the schistosomiasis control program started in 1950, and then we use a sample of people born between 1940 and 1955 to estimate the baseline model. If our findings are truly due to the real schistosomiasis control campaign, we also expect limited effects in this sample. As shown in Table A5, the results in these two columns are consistent with our expectation.

Finally, we use the empirical distribution of placebo coefficients to evaluate the results in DID studies. In particular, we randomly assign the pre-control infection rates to counties within each endemic province. Then we merge the fake disease data with the true individual-level census data and estimate equation (1) to obtain a placebo coefficient. Panel A of Figure A4 shows the distribution of 1,000 placebo coefficients in the female sample. For example, the distribution of placebo effects on junior high school completion is centered at 0.09, ranging from -0.05 to 0.24.[31] The vertical line representing our true DID estimate lies in the extreme right tail of the whole distribution, which implies that the empirical p-value is 0.007. Similarly, our baseline estimates of the effect on years of schooling and completing primary school are also statistically significant at the 1 percent level. In contrast, we find that a large proportion of placebo coefficients are larger than the baseline estimates in the subfigures of literacy and employment status. This echoes the insignificant results from the cohort analysis and the two-way fixed effects estimator with heterogeneous treatment effects. Panel B of Figure A4 plots the same empirical distribution of placebo coefficients in the male sample. It demonstrates that our baseline estimates of the effect on schooling years and completing junior high school are significant at the 1 percent level, and the estimate of the effect on completing primary school is significant at the 5 percent level. We cannot reject the null hypothesis that the schistosomiasis control program had no effect on literacy and employment status in their thirties. Taken together, this randomization inference method provides reinforcement that our main findings are not merely statistical artifacts.

---

[31] Schistosomiasis prevalence was clustered at the province level, and we only randomly assign infection rates within each province, so the distribution is not centered around 0. Exploiting the province-level variation can already reveal some positive effects of the deworming campaign, but not large enough to generate our baseline results.

IV Empirical Results from the CFPS Data

In this section, we employ the CFPS dataset to achieve three goals. First, we replicate what we have found in a smaller but finer sample. Second, we evaluate the long-term effects of early-life exposure to the schistosomiasis control campaign on adult economic status over 50 years later. Third, we link information across three generations and examine the intergenerational effects of the deworming campaign on educational attainment.

A Replication of the Baseline Results

We refine the baseline estimates from the census sample in three ways. First and foremost, we directly control for migration in our regressions. Since the CFPS does not provide detailed birthplace information, we drop people who moved out of their birthplaces (14.85% of the full sample). Table A6 shows that the resulting sample is still representative, as there are no statistically large differences in the outcomes of interest between movers and non-movers. Furthermore, we use the CFPS data to examine whether the disease or the disease control program had an impact on migration. As shown in Table A7, the pre-control schistosomiasis infection rate was unrelated to the possibility of moving out of birthplaces, and the deworming campaign had no effects on people's migration decisions. This implies that our estimates using the census sample do not suffer too much from our no-migration assumption. Second, the measure of early-life exposure to schistosomiasis is more precise. Because of the transmission feature of schistosomiasis, our goal is to keep people living in rural areas during their childhood in the analysis sample. When using the census data, we can only select people with rural *hukou* in 1990. Now we can select people with rural *hukou* when they were 3 years old, which should be a better proxy for childhood living area. Third, we further control for parents' education levels in our regressions.[32] Beach et al. (2022) show that it is important to control for family background characteristics when estimating the long-run effects of disease infections.

Table 8 presents the replication results. In this small but refined sample, we still find a large positive effect on educational attainment. According to column 1, a

---

[32] We do not add this control variable in the census sample, as doing so would shrink the sample size a lot and cause a selection problem. This is because we only have parents' education information for people who lived with their parents in the census sample. Table A8 shows that these people differed significantly from others along many dimensions. People living with their parents tend to be younger, more educated, and more likely to have a job.

reduction of 10 percentage points in infection rate led to an increase of 0.61 years in years of schooling for rural women and 0.48 years for rural men. Overall, the estimated effects in this table are larger than our baseline results using the census data, mainly because we further control for parents' educational levels. The omitted variable (i.e., parents' educational level) is positively correlated with the dependent variable but negatively correlated with the independent variable of interest, so the direction of the bias is negative.[33] Indeed, Table A9 presents smaller estimated effects when we do not control for parents' educational levels.

[Insert Table 8 here]

We use the census and the CFPS data to find that the schistosomiasis control campaign increased more schooling for females relative to males. This gender heterogeneity is consistent with the existing economic theory (Pitt, Rosenzweig and Hassan, 2012). The model assumes that investments in schooling and health and the choice of occupation are made in a Roy-type economy. Each occupation requires a bundle of brawn and skill. Schooling and health are complements in the production of skill, while only health contributes to the accumulation of brawn. Two key biological facts are embedded in the model: (1) men are stronger than women on average; and (2) health improvement increases brawn substantially more for men than for women.[34] Then in an economy where brawn is productive, health improvement brought by the disease control campaign will have a positive effect on education for females but an ambiguous effect for males.[35] The reduction in morbidity greatly augments the brawn of men and raise their productivity in occupations requiring raw labor, so staying in school is less profitable for men due to their higher opportunity cost of schooling. Given that the Chinese economy in our study period was quite underdeveloped, our paper provides reduced-form evidence consistent with this model. The model also implies that larger men may receive less schooling than smaller men. Because larger men have more brawn, they will have a higher opportunity cost of schooling. In

---

[33] As shown in Figure A2, there is a negative correlation between educational level and infection rate before the deworming campaign.

[34] These two biological facts have been documented in the medical literature (Mathiowetz et al., 1985; Round et al., 1999).

[35] This explains why empirical research finds mixed effects of public health interventions on educational attainment. We use the literature on malaria eradication as an example. Barreca (2010), Lucas (2010), and Barofsky, Anekwe and Chase (2015) show that malaria eradication significantly raised educational attainment in the United States, Sri Lanka, Paraguay, and southwestern Uganda. However, Cutler et al. (2010), Bleakley (2010), and Venkataramani (2012) find mixed effects in India and Mexico.

contrast, larger women are more likely to have higher levels of schooling than smaller women, as schooling and health are complements in the production of skill. Table A10 provides empirical evidence for this implication. We separate people into high and low endowments of brawn based on their adult heights. The positive effects of the deworming campaign on years of schooling concentrate on shorter men and taller women.

Column 5 of Table 8 reports the long-term effect of the deworming campaign on employment status in one's fifties. This is not a replication of our baseline result using the census data (as shown in column 5 of Table 3). For people born around the deworming campaign, though their educational attainment measured in 1990 and 2010 is almost the same, their employment status measured in two years is quite different. They were at their prime age in 1990, but many of them were retired or approaching retirement age in 2010.[36] We find that if a county reduced its pre-control schistosomiasis infection rate from 10 percent to zero, a rural woman born in this county would be about 7.53 percentage points more likely to have a job in 2010. This result suggests that the deworming campaign could have a long lasting effect on women's ability to work over 50 years later.

B Effects on Adult Economic Status

In this subsection, we examine the effects of early-life exposure to the schistosomiasis control program on adult economic status over 50 years later. We winsorize the outcomes of interest at the 1st and 99th percentile and then take the natural logarithm of them. Liu et al. (2008) point out that it is difficult to define individual income or separate individual contributions to household income in rural households in China, so we start by using per capita household income as the dependent variable in column 1 of Table 9. We find that if a man was born in a county that witnessed a 10 percent decline in the schistosomiasis infection rate, his household's per capita income would increase by 10.59 percent. This positive effect on the household income of treated women slightly decreased to 9.69 percent. In the next two columns, we follow Cutler et al. (2010) and explore the effects on household expenditure and consumption.[37] Based on our estimates, reducing the infection rate from 10 percent to zero would lead to a 8.77 (10.44) percent increase in per capita household expenditure

---

[36] The female employment rate is about 93% in the census sample, but it is only about 51% in the CFPS sample.

[37] Total expenditure is consumption plus transfer expenditure, welfare expenditure, and housing-related expenditure.

(consumption) for treated rural males, but we do not find the same effects for treated females.[38] The average household expenditure per capita in our sample was about 6,800 CNY, so in terms of money amount, this effect equals an increase of 596 CNY (roughly $88). In addition, the schistosomiasis control program improved households' asset accumulation. For treated men, the same decrease in schistosomiasis infection rate was associated with a 13.82 percent increase in per capita household net worth. In contrast, the effect for treated women was economically and statistically small.

[Insert Table 9 here]

Identifying mediators is a central challenge in studies of early life interventions, requiring either additional sources of exogenous variation in mediators or strong assumptions regarding the relationship between variables (Heckman, Pinto and Savelyev, 2013). The sequential ignorability condition assumed in some studies is not tenable here, as most outcomes are proxies for human capital and not measured in sequence throughout life. For this reason, we report the effects on potential mediators alongside the effects on the final outcome of interest. We first explore the effects of the deworming campaign on labor market outcomes. The results are reported in Table A11. In column 1, we find that the deworming campaign led to an increase of working hours for rural females. This effect mainly comes from the extensive margin of female employment, which is consistent with our result in column 5 of Table 8.[39] Column 2 shows that the deworming program helped rural men find more prestigious jobs, measured by the Standard International Occupational Prestige Scale (SIOPS).[40] As we do not find positive effect on employment status of rural males, our results imply that the positive labor market effects on males concentrate on the intensive margin of employment. We next estimate the effects on different measures of adult health and report the results in Table A12. Although we find no long-term effects on height, body mass index, or self-reported health, there is some evidence of persistent health gains in terms of interviewer-reported health for rural men.[41] To complete our preliminary analysis on the pathways between early-life health and adult income, we

---

[38] The same result pattern is also found in Cutler et al. (2010). They argue that the low female labor participation rate and the improvement in male productivity may explain this result.

[39] If we drop unemployed women in this regression, the DID coefficient becomes small and statistically insignificant.

[40] SIOPS is a scale that provides hierarchical ranking from the least to most esteemed occupations according to average societal ratings (Treiman, 1977).

[41] Self-reported health is a dummy variable that equals 1 if the respondent thinks (s)he is healthier than peers. Interviewer-reported health is a dummy variable that equals 1 if the interviewer thinks the respondent's health condition is above 5 on a scale of 1 to 7.

test if the positive relationship between these potential mechanisms and per capita household income exists in our sample. Table A13 shows that all the above relevant mechanisms have a strong positive correlation with adult income. To summarize, education, labor market success, and health condition all partially explain why the deworming campaign had a positive effect on adult economic status.

C Intergenerational Effects on Education

The detailed CFPS dataset offers us a rare opportunity to study the intergenerational effects of the disease control campaign on educational attainment. To avoid confusion, people born between 1940 and 1970 are referred to as Generation 1. We refer to their parents as Generation 0 and their children as Generation 2.

Schistosomiasis is a disease of poverty, given its close link with a safe and sanitary condition. Poor children in rural areas were highly exposed, whereas wealthier parents could establish a safer environment and purchase medical care for their children. Thus, children from poor families may have benefited more from the decrease in the prevalence of schistosomiasis. To analyse the differing impact of the deworming program, we separate Generation 1 into high and low socioeconomic backgrounds based on whether their fathers (i.e., Generation 0) were literate.[42] Table 10 displays the results from estimating equation (1) in four samples separated by gender and family background. Comparing panels A and B, we find that the estimated effects are larger for women from low socioeconomic background families. All the coefficients in panel A are more significant economically and statistically than those in panel B. As shown in panels C and D, we also find that rural men whose fathers were illiterate benefited more from the deworming campaign. From an intra-family perspective, compared with boys, girls often received less nutrition and financial support at that time in rural households. Thus, the heterogeneous education effect across gender is also consistent with this socioeconomic-dependent story. Taken together, these findings suggest that people from a low socioeconomic background benefited more from the schistosomiasis control program. The deworming campaign not only raised the average educational level in rural China, but also reduced educational inequality across gender and generations.

[Insert Table 10 here]

---

[42] About 33% of fathers were literate in the CFPS sample.

We next study the intergenerational effects of the schistosomiasis control campaign on the children of Generation 1 (i.e., Generation 2). Bütikofer and Salvanes (2020) discuss two possible mechanisms through which the benefits from a disease control program can transmit across generations. The first channel is through better living standards. We have provided some suggestive evidence of this channel in the last subsection. The second and more direct channel is through biological changes. As we lack data on birth information, this calls for further research. To address whether there is selection into the sample, we show in Table A14 that the schistosomiasis control program did not change people's fertility choice on the extensive or intensive margins.[43] Due to data limitation on the second generation, we focus on the educational attainment of Generation 1's first child who was at least 25 years old in 2010. Table 11 shows that Generation 2 had more years of schooling if their mothers experienced a larger drop in early-life exposure to schistosomiasis. However, treated fathers did not have similar intergenerational effects on the second generation. This result reflects the intra-family gender roles in traditional rural Chinese households. That is, mothers are homemakers, while fathers are breadwinners. Unlike Generation 1 who were more likely to complete primary and junior high school, Generation 2 were more likely to complete senior high school. The heterogeneity of the deworming effects at different educational levels is consistent with the rapid development in rural China.

[Insert Table 11 here]

V Conclusion

China used to be the largest endemic area of schistosomiasis in the 20th century. In the late 1950s, China's central government launched an unprecedented large-scale schistosomiasis control campaign to eradicate this disease. This paper exploits this deworming campaign to evaluate its long-term effects on human capital accumulation and economic status in adulthood. To conduct the analysis, we compile a county-level dataset on schistosomiasis infection rate and match it with the 1990 population census and the 2010 CFPS household survey. We employ a difference-in-differences specification, where we compare adult outcomes of people born before and after the deworming program across counties with varying pre-control schistosomiasis prevalence.

---

[43] About 98% of our sample had at least one child.

Using the cohort DID method, we find that the disease control campaign led to increased educational attainment among the rural population. Compared with the literature on malaria eradication, our estimates imply a larger education impact, consistent with the chronic and low-mortality features of schistosomiasis.[44] The education effect was larger on people from a lower socioeconomic background, such as girls and people whose fathers were not literate, suggesting that the deworming program reduced socioeconomic inequalities in adulthood. The deworming program also increased household income, consumption, and asset accumulation over 50 years later. We provide suggestive evidence that increased educational attainment, labor market success, and better adult health are potential mechanisms explaining the effect of the program on adult economic status. Future research can explore the relative weight of each possible channel for the increase in adult income. The effect of the deworming campaign persisted across generations, as we find a positive education effect on the children of the treated rural females.

Although we lack data to calculate a rate of return to the schistosomiasis control program, this paper presents clear policy implications for some countries that are still suffering from this disease. Many countries have launched preventive chemotherapy campaigns, but unfortunately just as many people suffer from schistosomiasis today as they did 50 years ago (Sokolow et al., 2018). Overall, our findings offer new justifications for mass deworming efforts, even in a very underdeveloped and unstable economy. Moreover, our results suggest that earlier deworming interventions may be more effective than the currently popular school-based interventions. With that said, more observational studies and clinical trials on the treatment timing are solicited to better reduce the disease burden of schistosomiasis.

---

[44] For instance, Lucas (2010) finds that reducing malaria incidence by 10 percentage points led to an increase in completed schooling of about 0.1 year and an increase in the probability of being literate by 1 to 2 percentage points for females in Paraguay and Sri Lanka.

References

Ajanga, Antony, Nicholas JS Lwambo, Lynsey Blair, Ursuline Nyandindi, Alan Fenwick and Simon Brooker. 2006. "Schistosoma mansoni in Pregnancy and Associations with Anaemia in Northwest Tanzania." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 100(1):59–63.

Almond, Douglas, Janet Currie and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56(4):1360–1446.

Almond, Douglas, Lena Edlund, Hongbin Li and Junsen Zhang. 2007. Long-Term Effects Of The 1959-1961 China Famine: Mainland China and Hong Kong. Working Paper 13384 National Bureau of Economic Research.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer and Edward Miguel. 2016. "Worms at Work: Long-Run Impacts of a Child Health Investment." *The Quarterly Journal of Economics* 131(4):1637–1680.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland and Michael Walton. 2017. "From proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31(4):73–102.

Barofsky, Jeremy, Tobenna D Anekwe and Claire Chase. 2015. "Malaria Eradication and Economic Outcomes in Sub-Saharan Africa: Evidence from Uganda." *Journal of Health Economics* 44:118–136.

Barreca, Alan I. 2010. "The Long-Term Economic Impact of In Utero and Postnatal Exposure to Malaria." *Journal of Human Resources* 45(4):865–892.

Beach, Brian, Ryan Brown, Joseph Ferrie, Martin Saavedra and Duncan Thomas. 2022. "Reevaluating the Long-Term Impact of In Utero Exposure to the 1918 Influenza Pandemic." *Journal of Political Economy* 130(7):000–000.

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119(1):249–275.

Bhalotra, Sonia, Martin Karlsson, Therese Nilsson and Nina Schwarz. 2022. "Infant Health, Cognitive Performance, and Earnings: Evidence from Inception of the Welfare State in Sweden." *Review of Economics and Statistics* 104(6):1138–1156.

Bharadwaj, Prashant, Katrine Vellesen Løken and Christopher Neilson. 2013. "Early Life Health Interventions and Academic Achievement." *American Economic Review* 103(5):1862–1891.

Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South." *The Quarterly Journal of Economics* 122(1):73–117.

Bleakley, Hoyt. 2010. "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics* 2(2):1–45.

Bütikofer, Aline and Kjell G Salvanes. 2020. "Disease Control and Inequality Reduction: Evidence from a Tuberculosis Testing and Vaccination Campaign." *The Review of Economic Studies* 87(5):2087–2125.

Callaway, Brantly and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225(2):200–230.

Chen, S-B, L Ai, W Hu, J Xu, R Bergquist, Z-Q Qin and J-H Chen. 2016. "New Anti-Schistosoma Approaches in The People's Republic of China: Development of Diagnostics, Vaccines and Other New Techniques Belonging to the 'Omics' Group." *Advances in Parasitology* 92:385–408.

Chen, Yi, Ziying Fan, Xiaomin Gu and Li-An Zhou. 2020. "Arrival of Young Talent: The Send-Down Movement and Rural Education in China." *American Economic Review* 110(11):3393–3430.

Chen, Yuyu and Li-An Zhou. 2007. "The Long-Term Health and Economic Consequences of the 1959–1961 Famine in China." *Journal of Health Economics* 26(4):659–681.

Colley, Daniel G, Amaya L Bustinduy, W Evan Secor and Charles H King. 2014. "Human Schistosomiasis." *The Lancet* 383(9936):2253–2264.

Croke, Kevin. 2014. The Long Run Effects of Early Childhood Deworming on Literacy and Numeracy: Evidence from Uganda. Unpublished manuscript.

Croke, Kevin and Rifat Atun. 2019. "The Long Run Impact of Early Childhood Deworming on Numeracy and Literacy: Evidence from Uganda." *PLoS Neglected Tropical Diseases* 13(1):e0007085.

Currie, Janet and Tom Vogl. 2013. "Early-Life Health and Adult Circumstance in Developing Countries." *Annual Review of Economics* 5(1):1–36.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal and Tom Vogl. 2010. "Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India." *American Economic Journal: Applied Economics* 2(2):72–94.

De Chaisemartin, Clément and Xavier d'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110(9):2964–96.

Dollar, David, Yiping Huang and Yang Yao. 2020. *China 2049: Economic Challenges of a Rising Global Power*. Washington, D.C.: Brookings Institution Press.

Ezeamama, Amara E, Amaya L Bustinduy, Allan K Nkwata, Leonardo Martinez, Noel Pabalan, Michael J Boivin and Charles H King. 2018. “Cognitive Deficits and Educational Loss in Children with Schistosome Infection—a Systematic Review and Meta-Analysis.” *PLoS Neglected Tropical Diseases* 12(1):e0005524.

Godwin, Fuseinisup, Edohsup Dominic, Gumah Kalifasup Bugre, Hamidsup Abdul-Wahab and Knight Dave. 2010. “Parasitic Infections and Anaemia During Pregnancy in the Kassena-Nankana District of Northern Ghana.” *Journal of Public Health and Epidemiology* 2(3):48–52.

Grönqvist, Hans, J Peter Nilsson and Per-Olof Robling. 2020. “Understanding How Low Levels of Early Lead Exposure Affect Children’s Life Trajectories.” *Journal of Political Economy* 128(9):3376–3433.

Heckman, James, Rodrigo Pinto and Peter Savelyev. 2013. “Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review* 103(6):2052–86.

Huang, Rongqing. 2009. “Assessing Accuracy in Age Reporting in China’s Population Census.” *Population Research* 33(6):30–41 (in Chinese).

Li, Nan and Xin Wei. 2017. “The Effect of Prevention and Cure of Schistosomiasis on Population Growth in China (1950–1990).” *Research in Chinses Economic History* 30(1):84–95 (in Chinese).

Li, Z-J, J Ge, J-R Dai, L-Y Wen, D-D Lin, H Madsen, X-N Zhou and S Lv. 2016. “Biology and Control of Snail Intermediate Host of Schistosoma japonicum in the People’s Republic of China.” *Advances in Parasitology* 92:197–236.

Liu, Gordon G, William H Dow, Alex Z Fu, John Akin and Peter Lance. 2008. “Income Productivity in China: On the Role of Health.” *Journal of Health Economics* 27(1):27–44.

Lucas, Adrienne M. 2010. “Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka.” *American Economic Journal: Applied Economics* 2(2):46–71.

Mao, Shoupai and Baoruo Shao. 1982. “Schistosomiasis Control in the People’s Republic of China.” *The American Journal of Tropical Medicine and Hygiene* 31(1):92–99.

Mathiowetz, Virgil, Nancy Kashman, Gloria Volland, Karen Weber, Mary Dowe, Sandra Rogers et al. 1985. “Grip and Pinch Strength: Normative Data for Adults.” *Archives of Physical Medicine and Rehabilitation* 66(2):69–74.

McDonald, Emily A, Sunthorn Pond-Tor, Blanca Jarilla, Marianne J Sagliba, Annaliza Gonzal, Amabelle J Amoylen, Remigio Olveda, Luz Acosta, Fusun Gundogan, Lisa M Ganley-Leal et al. 2014. "Schistosomiasis japonica During Pregnancy is Associated with Elevated Endotoxin Levels in Maternal and Placental Compartments." *The Journal of Infectious Diseases* 209(3):468–472.

Meng, Xin and Nancy Qian. 2009. The Long Term Consequences of Famine on Survivors: Evidence from a Unique Natural Experiment using China's Great Famine. Working Paper 14917 National Bureau of Economic Research.

Meng, Xin, Nancy Qian and Pierre Yared. 2015. "The Institutional Causes of China's Great Famine, 1959–1961." *The Review of Economic Studies* 82(4):1568–1611.

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

Naghavi, Mohsen, Amanuel Alemu Abajobir, Cristiana Abbafati, Kaja M Abbas, Foad Abd-Allah, Semaw Ferede Abera, Victor Aboyans, Olatunji Adetokunboh, Ashkan Afshin, Anurag Agrawal et al. 2017. "Global, Regional, and National Age-Sex Specific Mortality for 264 Causes of Death, 1980–2016: a Systematic Analysis for the Global Burden of Disease Study 2016." *The Lancet* 390(10100):1151–1210.

Nowakowski, Richard S. 2006. "Stable Neuron Numbers from Cradle to Grave." *Proceedings of the National Academy of Sciences* 103(33):12219–12220.

Ozier, Owen. 2018. "Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming." *American Economic Journal: Applied Economics* 10(3):235–62.

Pitt, Mark M, Mark R Rosenzweig and Mohammad Nazmul Hassan. 2012. "Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy." *American Economic Review* 102(7):3531–60.

Qian, Xinzhong. 1988. *China Schistosomiasis Atlas*. Shanghai: Chinese Map Publishing House.

Round, Joan M, David A Jones, JW Honour and Alan M Nevill. 1999. "Hormonal Factors in the Development of Differences in Strength between Boys and Girls during Adolescence: A Longitudinal Study." *Annals of Human Biology* 26(1):49–62.

Salawu, Oyetunde T and Alexander B Odaibo. 2014. "Maternal Schistosomiasis: a Growing Concern in sub-Saharan Africa." *Pathogens and Global Health* 108(6):263–270.

Sokolow, Susanne H, Chelsea L Wood, Isabel J Jones, Kevin D Lafferty, Armand M Kuris, Michael H Hsieh and Giulio A De Leo. 2018. "To Reduce the Global Burden of Human Schistosomiasis, Use 'Old Fashioned' Snail Control." *Trends in Parasitology* 34(1):23–40.

Sun, Liyang and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225(2):175–199.

Taylor-Robinson, David C, Nicola Maayan, Karla Soares-Weiser, Sarah Donegan and Paul Garner. 2015. "Deworming drugs for Soil-Transmitted Intestinal Worms in Children: Effects on Nutritional Indicators, Haemoglobin, and School Performance." *Cochrane Database of Systematic Reviews* 7:1–157.

Treiman, Donald. 1977. *Occupational Prestige in Comparative Perspective.* New York: Academic Press.

Venkataramani, Atheendar S. 2012. "Early Life Exposure to Malaria and Cognition in Adulthood: Evidence from Mexico." *Journal of Health Economics* 31(5):767–780.

Walder, Andrew G. 2014. "Rebellion and Repression in China, 1966–1971." *Social Science History* 38(3-4):513–539.

World Health Organization. 2016. "10 Facts about Schistosomiasis." http://www.who.int/features/factfiles/schistosomiasis/en/.

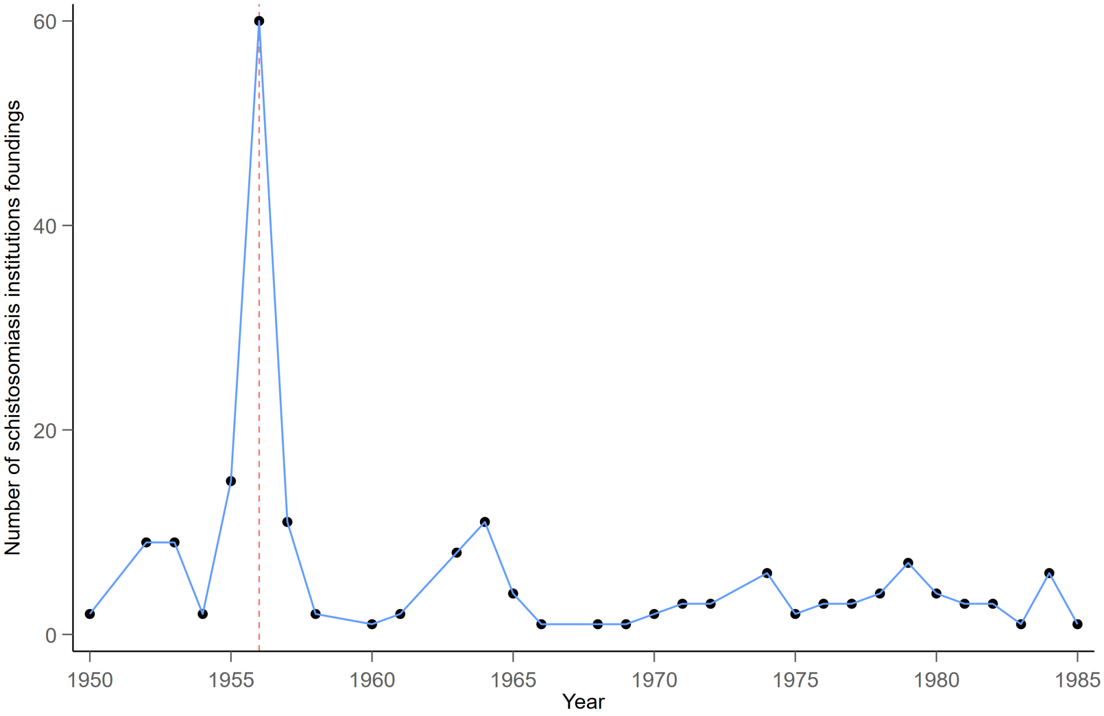
World Health Organization. 2017. "Schistosomiasis Fact Sheet." http://www.who.int/mediacentre/factsheets/fs115/en/.

Xu, J, P Steinman, D Maybe, X-N Zhou, S Lv, S-Z Li and R Peeling. 2016. "Evolution of the National Schistosomiasis Control Programs in the People's Republic of China." *Advances in Parasitology* 92:1–38.

Yuan, Hongchang, S-J Zhuo and S-J Zhang. 1989. "Epidemic Situation and Strategies for Schistosomiasis Control in Marshland and Lake Regions." *Chinese Journal of Schistosomiasis Control* 1:2–6 (in Chinese).

Zhao, Yaohui. 1997. "Labor Migration and Returns to Rural Education in China." *American Journal of Agricultural Economics* 79(4):1278–1287.

Figure 1. Time Trend in Founding of Medical Institutions Specialized in Schistosomiasis

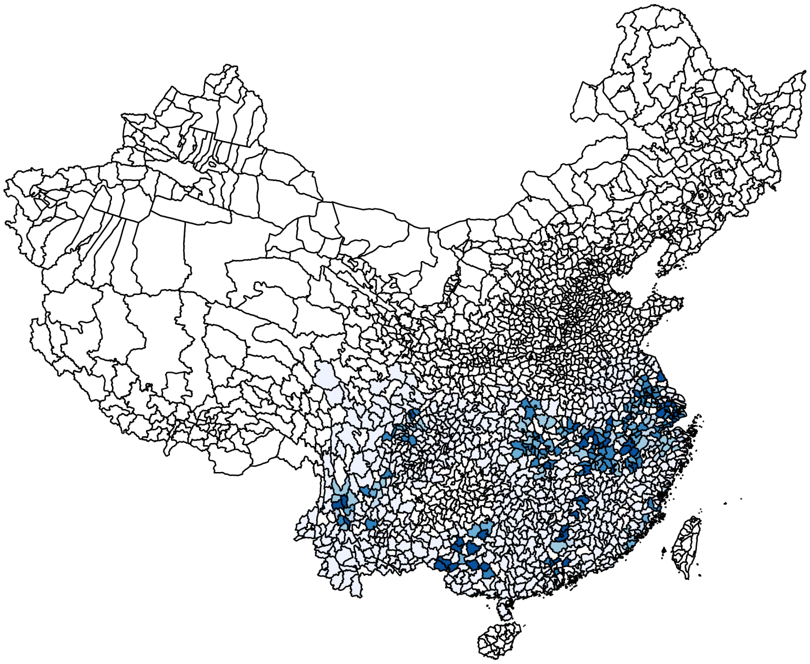


Source: Chinese County-Level Data on Hospitals and Epidemiology Stations, 1950-1985. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

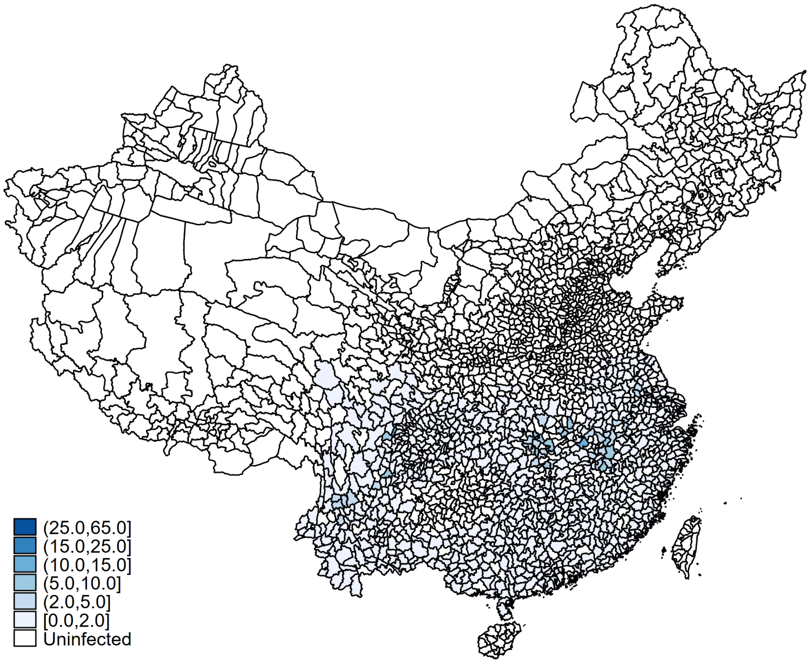
Notes: This figure plots the number of newly established medical institutions specialized in schistosomiasis from 1950 to 1985. The vertical line represents 1956.

Figure 2. Schistosomiasis Infection Rate

(a) Pre-control



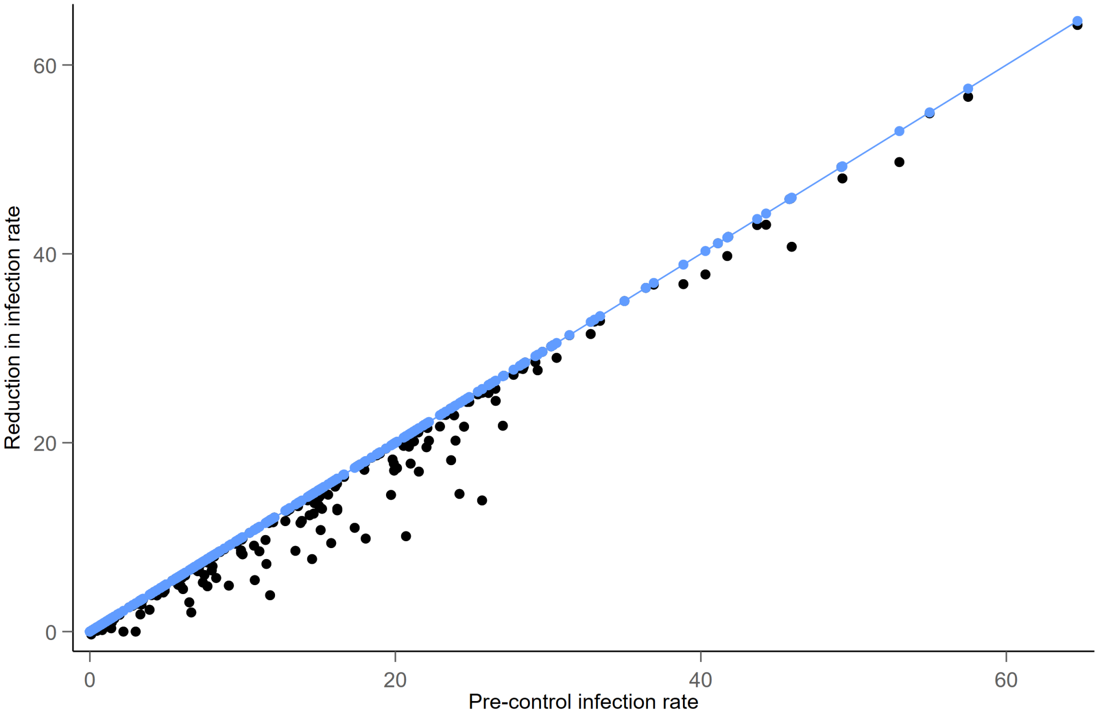
(b) Post-control



Source: China Schistosomiasis Atlas (Qian 1988).

Notes: Panels A and B of the map show the pre- and post-control schistosomiasis infection rate (the probability of an infection in a population) at the county level.

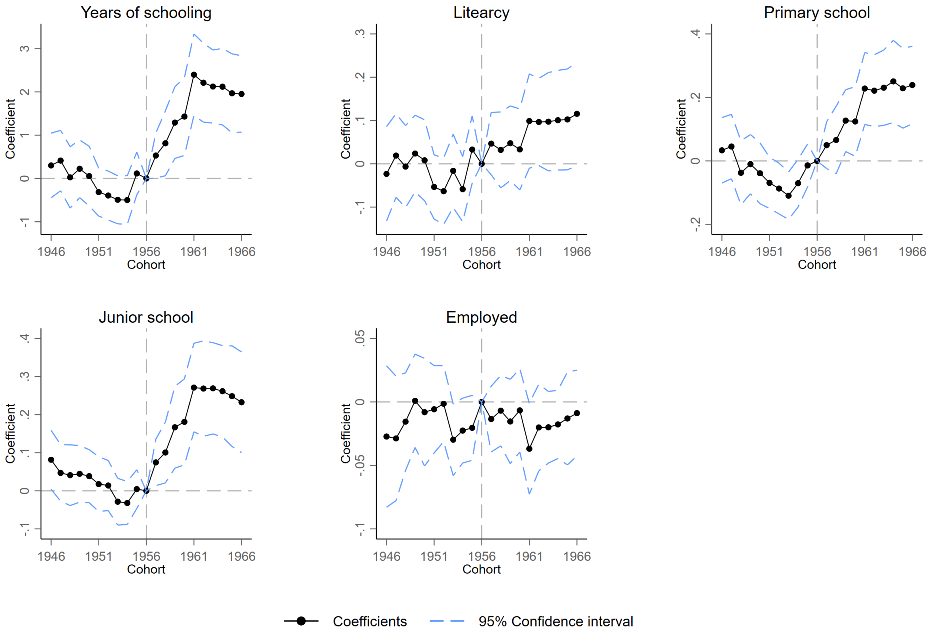
Figure 3. Pre-Control Infection Rate and Reduction in Infection Rate



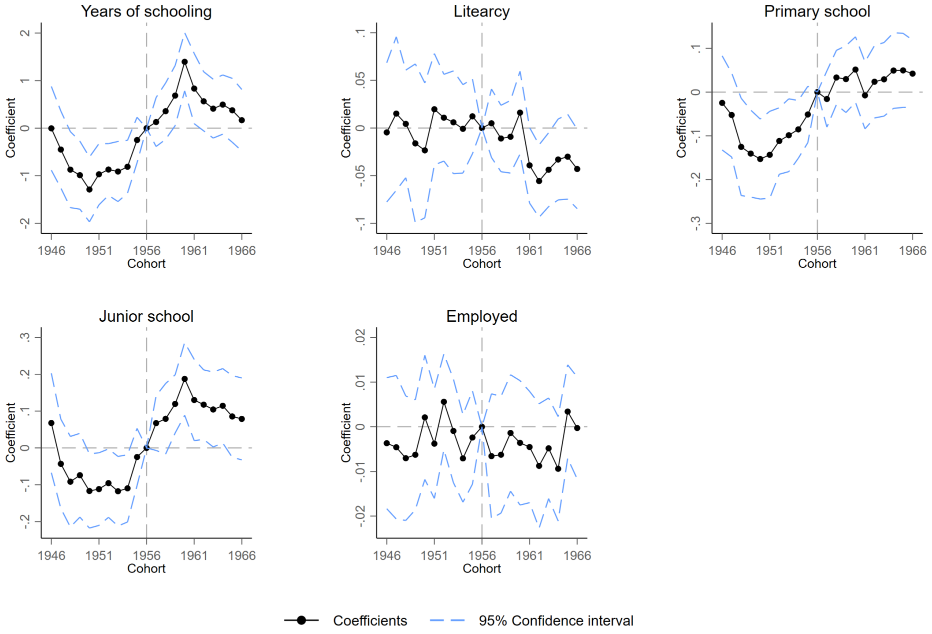
Notes: This figure plots the schistosomiasis control campaign effectiveness against initial disease prevalence at the county level. The y-axis is the reduction in infection rate. The x-axis is the pre-control infection rate. Each dot represents a county, and the dots on the diagonal are the counties that eliminated schistosomiasis.

Figure 4. Cohort Analysis

(a) Female



(b) Male



Notes: This figure plots the cohort-specific coefficients in equation (2). The effects on the cohort born in 1956 are set to zero. The outcomes are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior school, and being employed.

Table 1. Summary Statistics of the Census Sample

	Female	Male
Age	32.735	32.804
	(6.157)	(6.143)
Ethnicity (Han)	0.919	0.920
	(0.273)	(0.271)
Years of schooling	4.796	6.867
	(3.567)	(3.154)
Literacy	0.744	0.927
	(0.436)	(0.260)
Primary school	0.565	0.797
	(0.496)	(0.402)
Junior high school	0.228	0.423
	(0.420)	(0.494)
Employed	0.933	0.993
	(0.249)	(0.083)
Observations	651,812	676,545

Notes: The sample includes people with rural *hukou* and born between 1946 and 1966 in 12 endemic provinces (Anhui, Fujian, Guangdong, Guangxi, Hubei, Hunan, Jiangsu, Jiangxi, Shanghai, Sichuan (including Chongqing), Yunnan, and Zhejiang). All table entries represent sample means and standard deviations in parentheses. Ethnicity equals 1 if the individual is Han ethnicity. Literacy, primary school, junior high school, and employed are dummy variables indicating whether one is literate, completes primary school, completes junior high school, and has a job. We generate a continuous variable to measure years of schooling based on the original census data on education and completion status. The mapping is as follows: illiterate = 0 years, (in)complete primary school = 6 (3) years, (in)complete junior high school = 9 (7.5) years, (in)complete senior high school or specialized secondary school = 12 (10.5) years, and (in)complete college or junior college = 16 (14) years.

Table 2. Summary Statistics of the CFPS Sample

	Female	Male
Age	51.955	52.145
	(8.385)	(8.575)
Ethnicity (Han)	0.943	0.941
	(0.232)	(0.236)
Years of schooling	4.211	6.733
	(4.253)	(4.100)
Literacy	0.468	0.739
	(0.499)	(0.439)
Primary school	0.481	0.743
	(0.500)	(0.437)
Junior high school	0.247	0.461
	(0.431)	(0.498)
Employed	0.507	0.636
	(0.500)	(0.481)
per capita Household income	7,674	7,703
	(6,321)	(6,377)
per capita Expenditure	6,805	6,819
	(5,264)	(5,199)
per capita Consumption	5,602	5,691
	(4,259)	(4,287)
per capita Net worth	51,000	51,900
	(69,510)	(69,764)
Observations	4,850	4,889

Notes: The sample includes people with rural *hukou* at age 3 years and born between 1940 and 1970 in the 13 endemic provinces (Anhui, Chongqing, Fujian, Guangdong, Guangxi, Hubei, Hunan, Jiangsu, Jiangxi, Shanghai, Sichuan, Yunnan, and Zhejiang), plus four adjacent uninfected provinces (Gansu, Henan, Shaanxi, and Shandong). We exclude people who moved out of their birthplaces. All table entries represent sample means and standard deviations in parentheses, weighted by national sampling weights. Household income, expenditure, consumption, and net worth are measured in CNY in 2010.

Table 3. Baseline Results

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
	Panel A. Female				
sch × post	1.855	0.099	0.228	0.202	-0.002
	(0.348)	(0.045)	(0.044)	(0.056)	(0.012)
Observations	651,812	651,812	651,812	651,812	651,812
	Panel B. Male				
sch × post	1.137	-0.032	0.120	0.170	-0.002
	(0.246)	(0.024)	(0.038)	(0.036)	(0.003)
Observations	676,545	676,545	676,545	676,545	676,545
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports our baseline results from estimating equation (1). The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table 4. Controlling for Concurrent Events

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
			Panel A. Female		
sch × post	1.461	0.111	0.213	0.105	-0.010
	(0.384)	(0.049)	(0.050)	(0.063)	(0.015)
Observations	517,919	517,919	517,919	517,919	517,919
			Panel B. Male		
sch × post	1.246	0.004	0.157	0.164	-0.000
	(0.232)	(0.026)	(0.039)	(0.037)	(0.004)
Observations	539,500	539,500	539,500	539,500	539,500
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes
Control concurrent events	Yes	Yes	Yes	Yes	Yes

Notes: This table reports our results from estimating equation (3). The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. We control for the local intensity of the Great Famine, the Cultural Revolution, and the Send-Down Movement. Standard errors are reported in parentheses and are clustered at the county level.

Table 5. Reduced Form Instrument Variable Estimation

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
	Panel A. Female in infected areas				
water × post	2.151	0.155	0.262	0.202	-0.002
	(0.533)	(0.062)	(0.068)	(0.083)	(0.011)
Observations	551,273	551,273	551,273	551,273	551,273
	Panel B. Male in infected areas				
water × post	1.714	0.017	0.199	0.233	0.003
	(0.498)	(0.045)	(0.066)	(0.062)	(0.003)
Observations	573,097	573,097	573,097	573,097	573,097
	Panel C. Female in uninfected areas				
water × post	-0.295	0.009	-0.020	-0.052	0.064
	(0.531)	(0.118)	(0.111)	(0.126)	(0.052)
Observations	265,065	265,065	265,065	265,065	265,065
	Panel D. Male in uninfected areas				
water × post	-1.406	-0.065	-0.104	-0.196	-0.004
	(1.118)	(0.112)	(0.114)	(0.118)	(0.013)
Observations	268,284	268,284	268,284	268,284	268,284
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the reduced-form IV estimates in infected areas and uninfected areas (Henan, Shandong, Shanxi, and Gansu). $water_j$ is the proportion of water area in county j. The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table 6. Instrument Variable Estimation

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
	Panel A. Female				
sch × post	4.293	0.309	0.522	0.403	-0.004
	(1.180)	(0.109)	(0.124)	(0.191)	(0.022)
F statistic	21.564	21.564	21.564	21.564	21.564
Observations	551,273	551,273	551,273	551,273	551,273
	Panel B. Male				
sch × post	3.389	0.035	0.394	0.460	0.007
	(0.792)	(0.088)	(0.108)	(0.104)	(0.006)
F statistic	23.895	23.895	23.895	23.895	23.895
Observations	573,097	573,097	573,097	573,097	573,097
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports our IV results from estimating equation (5). We use as $water_j \times post_{ic}$ the instrument variable for $sch_j \times post_{ic}$. We use a sample consisting of marshland, lake regions, and plain regions with waterway networks. The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table 7. Critical Period of Deworming

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high
	Panel A. Female			
sch × cohort1	1.757	0.125	0.213	0.166
	(0.419)	(0.053)	(0.051)	(0.062)
sch × cohort2	-0.272	0.018	-0.036	-0.061
	(0.213)	(0.033)	(0.029)	(0.022)
Observations	723,278	723,278	723,278	723,278
	Panel B. Male			
sch × cohort1	1.059	-0.030	0.124	0.135
	(0.291)	(0.034)	(0.043)	(0.041)
sch × cohort2	0.026	0.006	0.024	-0.038
	(0.275)	(0.022)	(0.029)	(0.044)
Observations	751,627	751,627	751,627	751,627
County FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes

Notes: This table reports the results from estimating equation (6). $cohort^1_{ic}$ equals 1 if individual i was born between 1957 and 1966, and $cohort^2_{ic}$ equals 1 if individual i was born between 1952 and 1956. People born between 1942 and 1951 are left as the reference group. The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table 8. Baseline Results from the CFPS Sample

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
			Panel A. Female		
sch × post	6.137	0.606	0.942	0.510	0.753
	(2.327)	(0.312)	(0.266)	(0.238)	(0.277)
Observations	3,677	3,677	3,677	3,677	3,677
			Panel B. Male		
sch × post	4.761	0.405	0.417	0.501	-0.163
	(2.460)	(0.340)	(0.302)	(0.276)	(0.344)
Observations	3,804	3,804	3,804	3,804	3,804
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes

Notes: This table reports our baseline results using the CFPS sample from estimating equation (1). The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table 9. Effects on Adult Economic Status

	(1) Household income	(2) Expenditure	(3) Consumption	(4) Net worth
	Panel A. Female			
sch × post	0.969	0.498	0.360	0.116
	(0.452)	(0.382)	(0.358)	(0.608)
Observations	2,951	2,951	2,951	2,951
	Panel B. Male			
sch × post	1.059	0.877	1.044	1.382
	(0.411)	(0.338)	(0.389)	(0.641)
Observations	3,095	3,095	3,095	3,095
County FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated effects on adult economic status using the CFPS sample from estimating equation (1). The dependent variables are the log form of per capita household income, expenditure, consumption, and net worth. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table 10. Effects on Educational Attainment by Socioeconomic Background

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high
	Panel A. Female whose father is illiterate			
sch × post	6.212	0.720	0.906	0.695
	(3.123)	(0.346)	(0.349)	(0.304)
Observations	2,485	2,485	2,485	2,485
	Panel B. Female whose father is literate			
sch × post	0.252	0.074	0.162	-0.128
	(3.497)	(0.375)	(0.452)	(0.353)
Observations	1,672	1,672	1,672	1,672
	Panel C. Male whose father is illiterate			
sch × post	4.980	0.398	0.377	0.436
	(2.159)	(0.351)	(0.269)	(0.329)
Observations	2,556	2,556	2,556	2,556
	Panel D. Male whose father is literate			
sch × post	0.951	0.042	0.058	0.166
	(4.122)	(0.395)	(0.465)	(0.403)
Observations	1,643	1,643	1,643	1,643
County FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes

Notes: This table reports heterogeneous effects on educational attainment by socioeconomic background. The dependent variables are years of schooling and three dummy variables indicating literacy, completing primary school, and completing junior high school. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and mothers' educational level. Standard errors are reported in parentheses and are clustered at the county level.

Table 11. Effects on Educational Attainment of the Second Generation

Dependent variable: Generation 2	(1) Schooling (years)	(2) Primary school	(3) Junior high	(4) Senior high	(5) College
	Panel A. Generation 1 Female				
sch × post	6.620	0.261	0.182	0.865	0.349
	(2.432)	(0.222)	(0.312)	(0.396)	(0.273)
Observations	2,090	2,090	2,090	2,090	2,090
	Panel B. Generation 1 Male				
sch × post	0.595	-0.258	-0.622	0.305	0.203
	(5.932)	(0.366)	(0.681)	(0.511)	(0.447)
Observations	1,980	1,980	1,980	1,980	1,980
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes

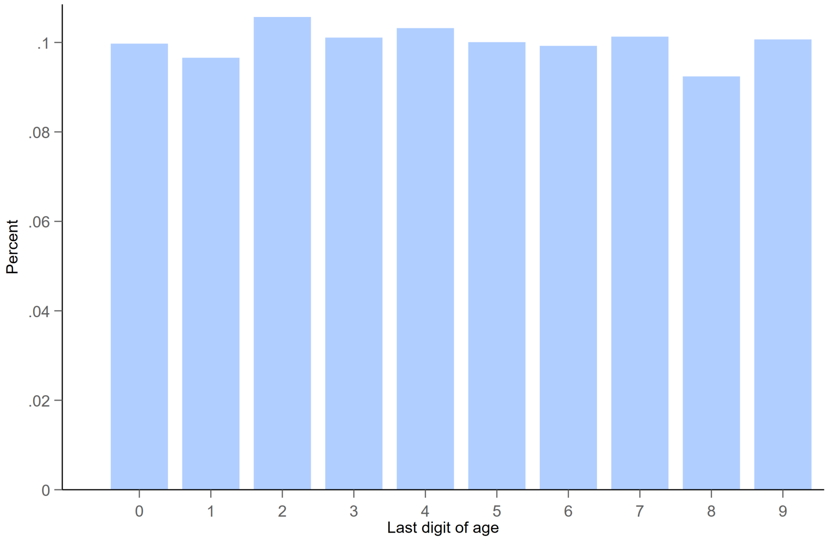
Notes: This table reports the estimated effects on educational attainment of the second generation. The dependent variables are the second generation's years of schooling and four dummy variables indicating completing primary school, completing junior high school, completing senior high school, and completing college. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity, the second generation's gender and age, and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Online Appendix

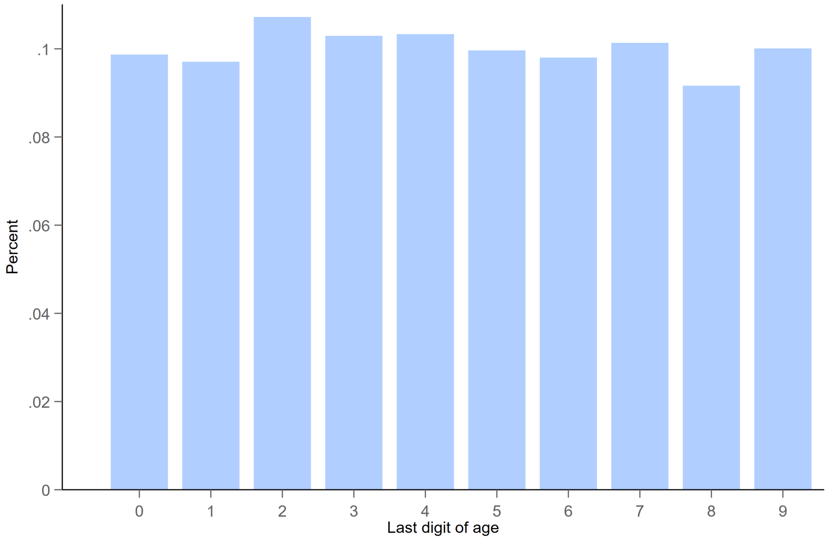
A Appendix Figures and Tables

Figure A1. Age Heaping

(a) Female



(b) Male



Notes: This figure plots the distribution of the last digit of age by gender among rural people between 19 and 58 years old in the 1 percent sample of the 1990 China Population Census.

Figure A2. Time Trend in Outcomes

(a) Female

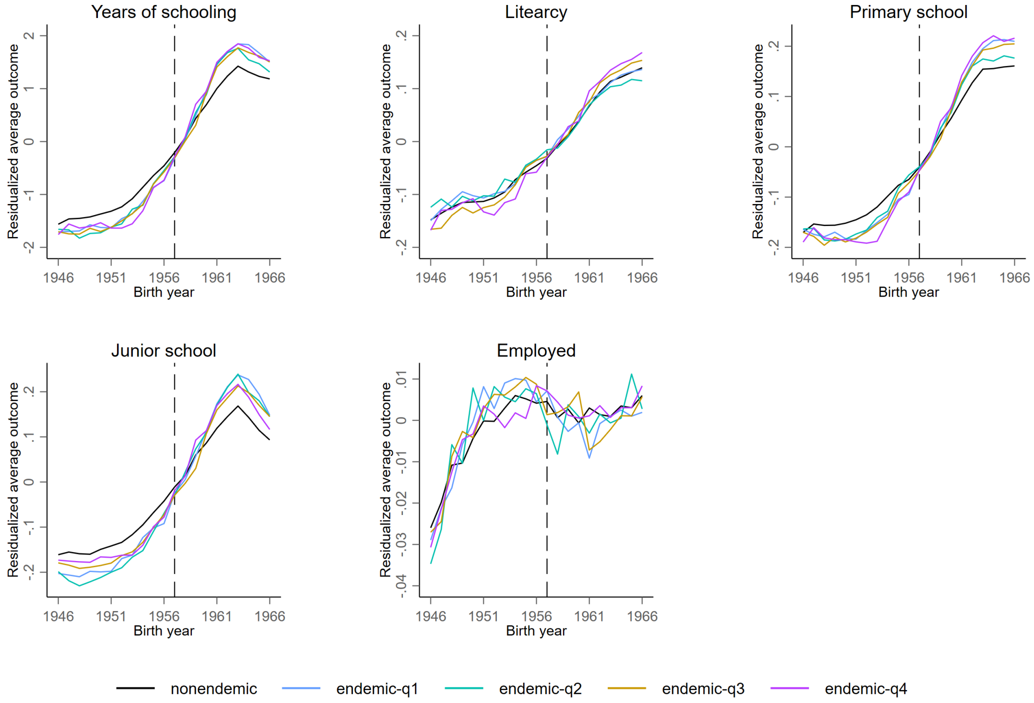
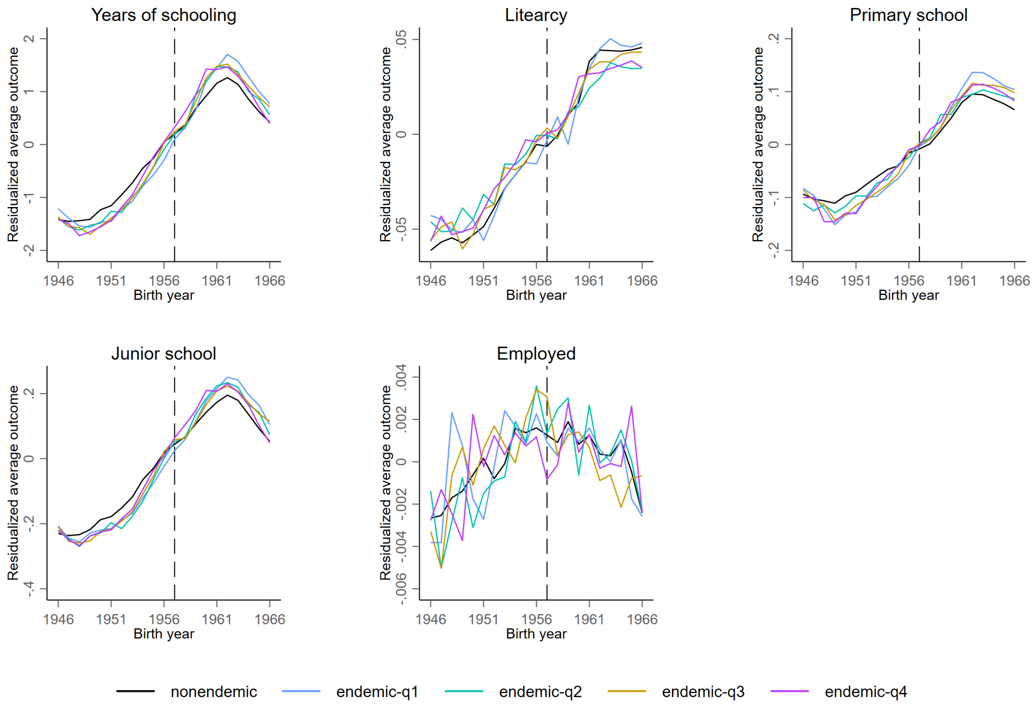


Figure A2. Time Trend in Outcomes (cont.)

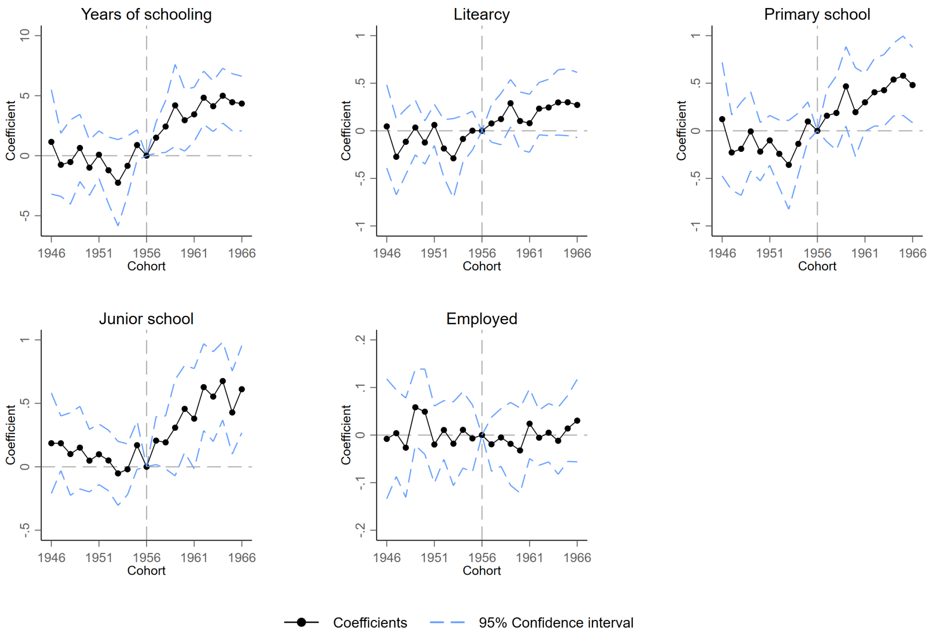
(b) Male



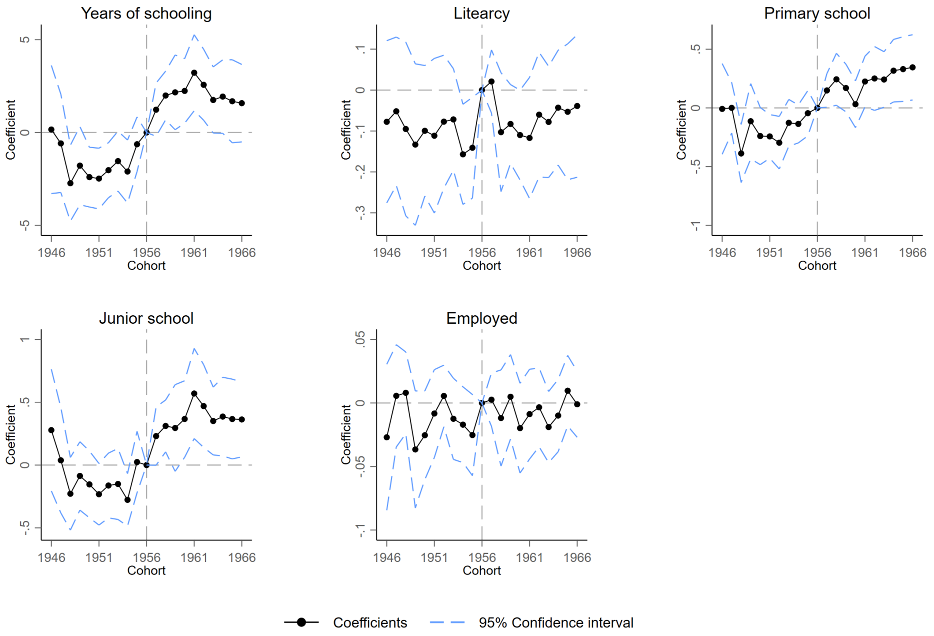
Notes: The figure plots the cohort-level mean residual outcomes conditional on ethnicity and county fixed effects for rural women (panel A) and men (panel B) born from 1946 to 1966. The outcomes are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. Different colored lines denote the average outcomes of counties in four quartiles of the pre-control infection rate. The black line denotes the average outcomes in uninfected counties. The vertical dashed line marks the 1957 cohort.

Figure A3. Cohort Analysis (IV)

(a) Female



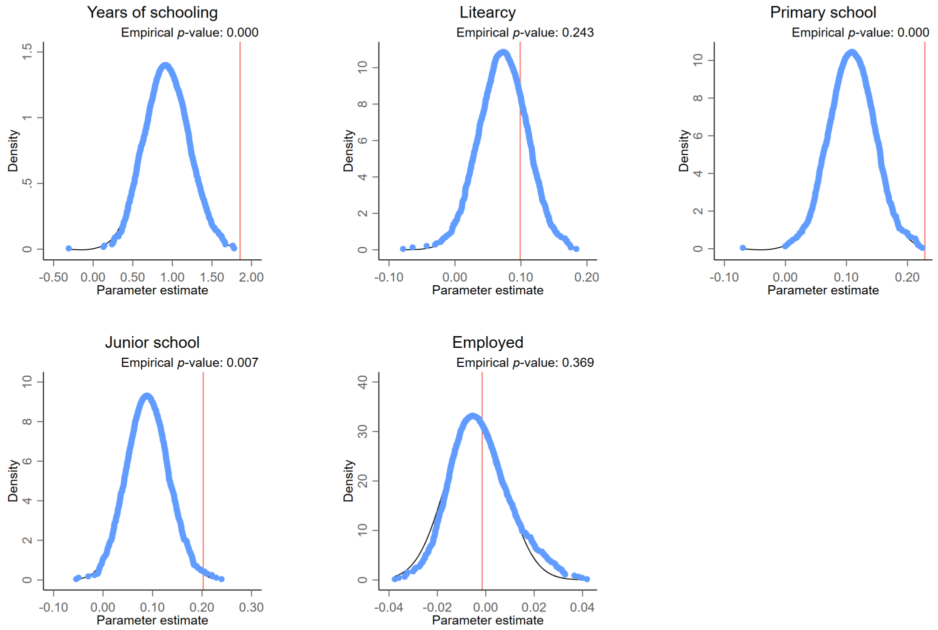
(b) Male



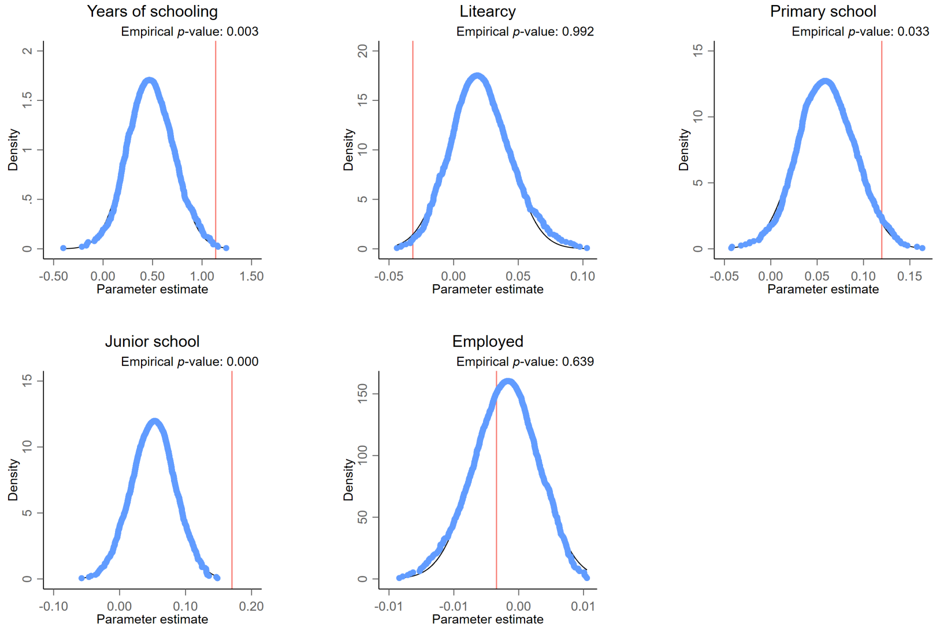
Notes: This figure plots the cohort-specific coefficients in equation (2) with an instrument variable for pre-control infection rate. The effects on the cohort born in 1956 are set to zero. The outcomes are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed.

Figure A4. Randomization Inference

(a) Female



(b) Male



Notes: This figure plots the distribution of 1000 placebo DID coefficients. The vertical lines represent our baseline results in Table 3. We randomly assign the county-level pre-control schistosomiasis infection rates within each province. The outcomes are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed.

Table A1. Correlation between Pre-Control Infection Rate and the Great Leap Forward Intensity

	(1) sch	(2) Steel growth	(3) Communal dining	(4) Anti-right	(5) PCA
sch	1				
Steel growth	-0.165	1			
	(0.628)				
Communal dining	0.312	-0.168	1		
	(0.350)	(0.621)			
Anti-right	0.407	-0.381	0.658	1	
	(0.214)	(0.248)	(0.028)		
PCA	0.393	-0.577	0.827	0.908	1
	(0.232)	(0.063)	(0.002)	(0.000)	

Notes: Column 1 presents that the pre-control infection rate was uncorrelated with any measure of political zealousness at the province level. The Great Leap Forward intensity measures come from Meng, Qian and Yared (2015). They are per capita steel production growth (avg. 1958-1960), the participation rate in communal dining halls (avg. 1955-1958), the magnitude of the 1957 Anti-Right purges, and the first principal component of the above three variables. Significance levels are reported in parentheses.

Table A2. Effects on Number of Schools

	(1) #Primary schools	(2) #Secondary schools
sch × post	-0.589	-0.003
	(0.229)	(0.006)
Observations	6,497	5,825
County FEs	Yes	Yes
Year FEs	Yes	Yes

Notes: This table reports the estimated results by estimating $y_{jt} = \alpha + \beta(sch_j \times post_t) + \delta_t + \delta_j + \varepsilon_{jt}$. We use county-level data between 1950 and 1966 from Chen et al. (2020). The dependent variables are the number of primary schools per 100,000 people in column 1 and the number of secondary schools per 100,000 people in column 2, scaled by population in 1964. We add county fixed effects and year fixed effects in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table A3. First Stage of IV Regression

	(1) Female	(2) Male
water × post	0.501	0.506
	(0.108)	(0.103)
Observations	551,273	573,097
County FEs	Yes	Yes
Cohort FEs	Yes	Yes

Notes: This table reports the first stage of IV regression from estimating equation (4). The dependent variable is $D_{ijc} = sch_j \times post_{ic}$. Columns 1 and 2 use female and male sample, respectively. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table A4. Critical Period of Deworming

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high
		Panel A. Female		
sch × cohort1	2.215	0.272	0.218	0.162
	(0.555)	(0.059)	(0.067)	(0.069)
sch × cohort2	0.194	0.166	-0.030	-0.063
	(0.385)	(0.060)	(0.059)	(0.028)
sch × cohort3	0.467	0.146	0.006	-0.001
	(0.260)	(0.044)	(0.042)	(0.014)
Observations	836,440	836,440	836,440	836,440
		Panel B. Male		
sch × cohort1	1.674	0.019	0.254	0.168
	(0.387)	(0.054)	(0.058)	(0.050)
sch × cohort2	0.645	0.055	0.155	-0.005
	(0.355)	(0.043)	(0.049)	(0.053)
sch × cohort3	0.618	0.049	0.131	0.033
	(0.219)	(0.029)	(0.038)	(0.027)
Observations	873,635	873,635	873,635	873,635
County FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes

Notes: This table reports the heterogeneous effects on educational attainment across the age of receiving deworming treatment. $cohort^1_{ic}$ equals 1 if individual i was born between 1957 and 1966, $cohort^2_{ic}$ equals 1 if individual i was born between 1952 and 1956, and $cohort^3_{ic}$ equals 1 if individual i was born between 1942 and 1951. People born between 1935 and 1941 are left as the reference group. The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table A5. Robustness Checks

Dependent variable	(1) Linear trends	(2) Mean reversion	(3) DID_m	(4) Urban sample	(5) Fake event
		Panel A. Female			
Schooling (years)	1.324	1.864	3.584	-0.279	-0.272
	(0.274)	(0.346)	(0.595)	(0.434)	(0.220)
Literacy	0.061	0.113	0.036	-0.031	0.041
	(0.030)	(0.034)	(0.060)	(0.030)	(0.037)
Primary school	0.144	0.210	0.387	-0.025	-0.050
	(0.034)	(0.043)	(0.073)	(0.045)	(0.032)
Junior high school	0.150	0.229	0.551	-0.033	-0.066
	(0.035)	(0.046)	(0.098)	(0.071)	(0.019)
Employed	-0.005	0.001	-0.025	-0.025	0.017
	(0.012)	(0.011)	(0.020)	(0.033)	(0.021)
Observations	651,812	651,812	350,149	82,317	376,802
		Panel B. Male			
Schooling (years)	1.100	1.022	2.786	0.063	-0.187
	(0.278)	(0.209)	(0.401)	(0.477)	(0.267)
Literacy	-0.011	-0.017	-0.049	0.049	0.024
	(0.013)	(0.010)	(0.034)	(0.022)	(0.021)
Primary school	0.068	0.064	0.260	0.036	0.007
	(0.031)	(0.022)	(0.034)	(0.029)	(0.031)
Junior high school	0.187	0.165	0.445	-0.043	-0.077
	(0.043)	(0.036)	(0.067)	(0.081)	(0.042)
Employed	-0.007	-0.002	-0.002	-0.004	0.003
	(0.005)	(0.003)	(0.004)	(0.024)	(0.003)
Observations	676,545	676,545	361,523	117,743	394,620

Notes: This table reports robustness checks of our baseline results. Column 1 includes county-specific linear time trends. In column 2, we include a series of interaction terms between the average of the dependent variable among people born from 1946 to 1956 and cohort dummies (i.e., $\bar{y}_{j,46-56} \times cohort_{ic}$). In column 3, we estimate a two-way fixed effects model with heterogeneous treatment effects, as described in De Chaisemartin and d'Haultfoeuille (2020). Columns 4 and 5 are falsification tests using an urban sample and a fake intervention in 1950. The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level in columns 1-2 and 4-5. Standard errors are clustered at the county level and computed by 1,000 bootstrap replications in column 3.

Table A6. Difference between Non-Movers and Movers

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
	Panel A. Female				
native	0.028	0.011	0.028	-0.021	0.028
	(0.299)	(0.033)	(0.035)	(0.032)	(0.044)
Observations	4,314	4,314	4,314	4,314	4,314
	Panel B. Male				
native	-0.456	-0.003	-0.019	-0.049	0.004
	(0.334)	(0.034)	(0.029)	(0.042)	(0.053)
Observations	4,313	4,313	4,313	4,313	4,313
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the comparison of educational attainment and employment status between movers and non-movers in the CFPS sample. The independent variable is a dummy variable indicating that the person did not move out of the birthplace. The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls are ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table A7. Effects on Migration

Dependent variable: Migration	(1)	(2)	(3)	(4)
	Panel A. Female			
sch	-0.007	-0.007		
	(0.004)	(0.005)		
sch × post			-0.207	0.069
			(0.210)	(0.246)
Observations	5,715	4,316	5,712	4,314
	Panel B. Male			
sch	-0.005	-0.004		
	(0.003)	(0.003)		
sch × post			0.209	0.013
			(0.235)	(0.298)
Observations	5,560	4,313	5,559	4,313
County FEs	No	No	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	Yes

Notes: This table presents the estimates of the effect of schistosomiasis prevalence and the deworming campaign on migration. The dependent variable is a dummy variable indicating that the person moved out of his or her birthplace. All regressions are weighted by national sampling weights. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table A8. Differences Between People Who Lived With Parents and People Who Did Not

	Yes	No	Difference
Age	30.265	33.600	-3.335
	(5.892)	(6.007)	(0.012)
Male	0.710	0.443	0.267
	(0.454)	(0.497)	(0.001)
Ethnicity (Han)	0.906	0.924	-0.019
	(0.292)	(0.264	(0.001))
Years of schooling	6.727	5.560	1.167
	(3.227)	(3.563)	(0.007)
Literate	0.911	0.813	0.098
	(0.284)	(0.390)	(0.001)
Primary school	0.781	0.650	0.131
	(0.413)	(0.477)	(0.001)
Junior high school	0.419	0.297	0.122
	(0.493)	(0.457)	(0.001)
Employed	0.978	0.959	0.019
	(0.147)	(0.198)	(0.000)
Observations	330,497	997,860	

Notes: See the note of Table 1 for the description of the sample and the definitions of variables. Standard deviations are in parentheses in columns 1 and 2. Column 3 reports the differences between people who lived with parents and who did not, which is obtained from a regression of demographic characteristics on an indicator for living with parents. Standard errors are in parentheses in column 3.

Table A9. Baseline Results from the CFPS Sample without Controlling Parents' Education

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
	Panel A. Female				
sch × post	4.154	0.424	0.575	0.418	0.609
	(2.537)	(0.268)	(0.302)	(0.236)	(0.387)
Observations	4,849	4,849	4,849	4,849	4,849
	Panel B. Male				
sch × post	1.480	0.014	0.045	0.123	0.133
	(2.282)	(0.253)	(0.235)	(0.293)	(0.292)
Observations	4,889	4,889	4,889	4,889	4,889
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports our baseline results using the CFPS sample from estimating equation (1). The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are reported in parentheses and are clustered at the county level.

Table A10. Heterogeneous Effects on Years of Schooling by Brawn Endowment

	(1) Female≤158cm	(2) Female≥158cm	(3) Male≤168cm	(4) Male≥168cm
sch × post	2.836	7.294	6.368	2.591
	(2.976)	(2.507)	(3.410)	(2.426)
Observations	1,996	2,035	2,050	2,169
County FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes

Notes: This table reports heterogeneous effects on years of schooling across brawn endowment. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls are ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table A11. Effects on Labor Market Outcomes

	(1) Working hours	(2) Occupational Prestige
	Panel A. Female	
sch × post	6.169	-3.690
	(2.260)	(10.917)
Observations	3,677	1,803
	Panel B. Male	
sch × post	1.133	19.246
	(3.872)	(9.227)
Observations	3,804	2,315
County FEs	Yes	Yes
Cohort FEs	Yes	Yes
Individual controls	Yes	Yes

Notes: This table reports the estimated effects on labor market outcomes. The dependent variables are working hours and Standard International Occupational Prestige Scale. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table A12. Effects on Adult Health

	(1) Height	(2) BMI	(3) Self-reported health	(4) Interviewer -reported health
			Panel A. Female	
sch × post	-0.656	-3.787	-0.041	0.338
	(4.972)	(2.303)	(0.370)	(0.303)
Observations	3,464	3,431	3,677	3,677
			Panel B. Male	
sch × post	-0.568	1.765	0.278	0.686
	(3.061)	(3.130)	(0.217)	(0.346)
Observations	3,748	3,737	3,804	3,804
County FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes

Notes: This table reports the estimated effects on adult health. The dependent variables are height, body mass index, self-reported health, and interviewer-reported health. Self-reported health is a dummy variable that equals 1 if the respondent thinks (s)he is healthier than peers. Interviewer-reported health is a dummy variable that equals 1 if the interviewer thinks the respondent's health condition is above 5 on a scale of 1 to 7. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table A13. Effects of Potential Mediators on Household Income

Mediators	(1)	(2)	(3)	(4)	(5)
Education	0.027		0.035		
	(0.005)		(0.006)		
Employed		0.144			
		(0.044)			
Occupational Prestige				0.007	
				(0.002)	
Health					0.084
					(0.017)
Observations	2,951	2,951	3,095	1,894	3,095
Sample	Female	Female	Male	Male	Male
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the relationship between potential mediators and adult economic status. The dependent variable is the log of per capita household income. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

Table A14. Effects on Fertility

	(1) $\mathbb{1}$(has children)	(2) No. children
	Panel A. Female	
sch × post	0.074	0.644
	(0.087)	(0.502)
Observations	3,677	3,677
	Panel B. Male	
sch × post	0.106	0.187
	(0.115)	(0.534)
Observations	3,804	3,804
County FEs	Yes	Yes
Cohort FEs	Yes	Yes
Individual controls	Yes	Yes

Notes: This table reports the estimated effects on fertility. The dependent variables are a dummy variable indicating if the person had a child and the number of children. All regressions are weighted by national sampling weights. We add county fixed effects and cohort fixed effects in all regressions. Individual controls include ethnicity and parents' educational levels. Standard errors are reported in parentheses and are clustered at the county level.

B Data: China Schistosomiasis Atlas

We describe the data collection process as follows. According to a unified statistical table and method, the county-level offices of schistosomiasis control collected the number of cases and reported it to the higher authority at the province level. Then the province-level offices reviewed and summarized the data and sent them to the central office for schistosomiasis control in Beijing. The central office released the final statistics after some corrections.

For each endemic province, the atlas contains a map (e.g., Figure B1) and a statistical table (e.g., Figure B2). The map plots the area of snail habitats and population infection rates. The table reports (from the first column to the last):

1. county name,
2. the number of endemic people's communes,[1]
3. the number of endemic production teams,[2]
4. population in endemic production teams,
5. the number of total patients,
6. the number of total chronic patients,
7. the number of current patients,
8. the number of current chronic patients,
9. pre-control infection rate,
10. current infection rate,
11. total area of snail habitats,
12. current area of snail habitats,
13. total and current area of snail habitats of three types.[3]

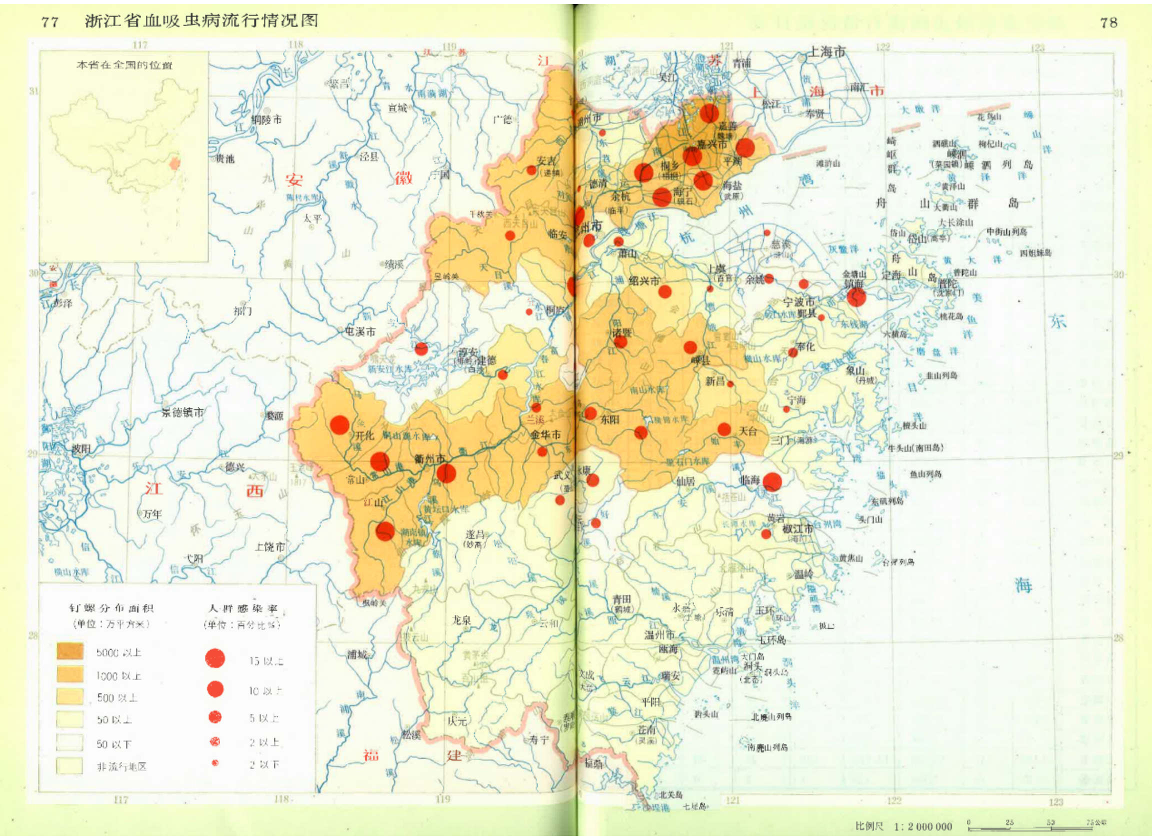
---

[1] The people's commune was the largest collective units in rural areas from 1958 to 1983. The communes had governmental, political, and economic functions during the Cultural Revolution.

[2] A production team was formerly the basic accounting and farm production unit in the people's commune system from 1958 to 1984.

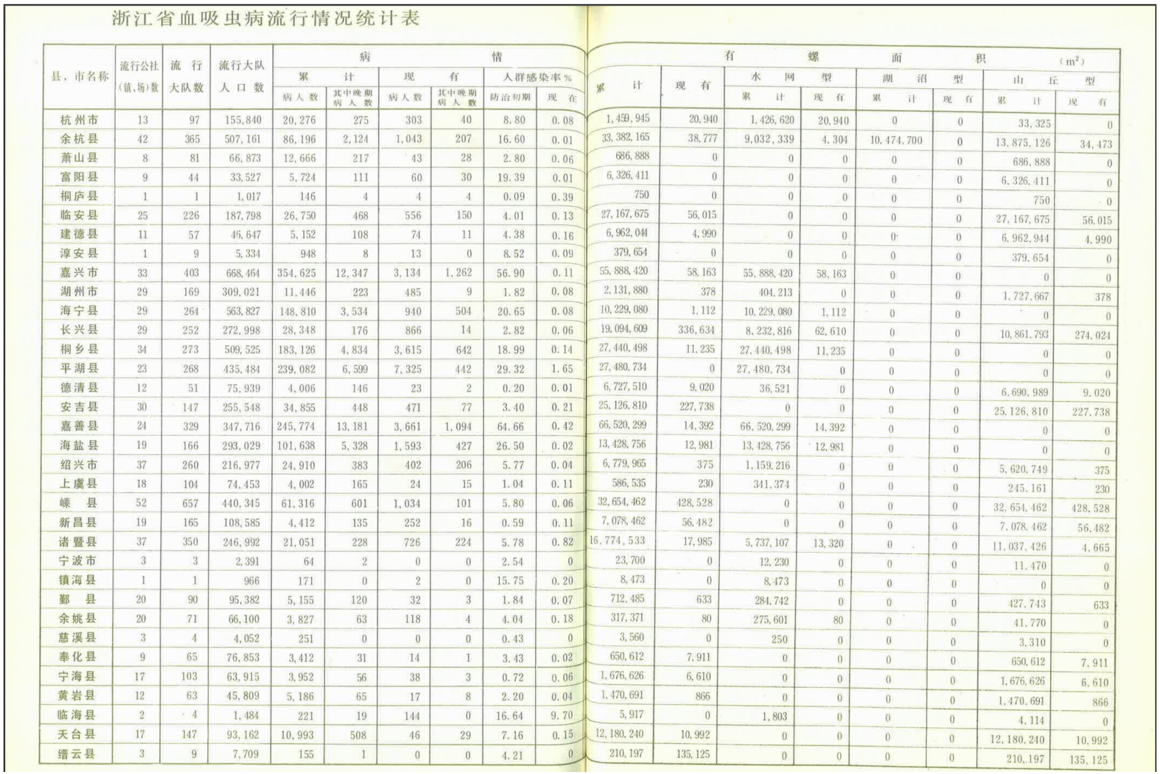
[3] Three types are plain regions with waterway networks, marshland and lake regions, hilly and mountainous regions

Figure B1. Schistosomiasis Prevalence Map of Zhejiang Province



Notes: The figure is a scanned schistosomiasis prevalence map of Zhejiang province from the China Schistosomiasis Atlas (Qian 1988).

Figure B2. Schistosomiasis Prevalence Statistical Table of Zhejiang Province



浙江省血吸虫病流行情况统计表

县、市名称	流行公社（镇、场）数	流行大队数	流行大队人口数	病情						有螺面积（m²）							
				累计		现有		人群感染率%		累计	现有	水网型		湖沼型		山丘型	
				病人数	其中晚期病人数	病人数	其中晚期病人数	防治初期	现在			累计	现有	累计	现有	累计	现有
杭州市	13	97	155,840	20,276	275	303	40	8.80	0.08	1,459,945	20,940	1,426,620	20,940	0	0	33,325	0
余杭县	42	365	507,161	86,196	2,124	1,043	207	16.60	0.01	33,382,165	38,777	9,032,339	4,304	10,474,700	0	13,875,126	34,473
萧山县	8	81	66,873	12,666	217	43	28	2.80	0.06	686,888	0	0	0	0	0	686,888	0
富阳县	9	44	33,527	5,724	111	60	30	19.39	0.01	6,326,411	0	0	0	0	0	6,326,411	0
桐庐县	1	1	1,017	146	4	4	4	0.09	0.39	750	0	0	0	0	0	750	0
临安县	25	226	187,798	26,750	468	556	150	4.01	0.13	27,167,675	56,015	0	0	0	0	27,167,675	56,015
建德县	11	57	46,647	5,152	108	74	11	4.38	0.16	6,962,044	4,990	0	0	0	0	6,962,944	4,990
淳安县	1	9	5,334	948	8	13	0	8.52	0.09	379,654	0	0	0	0	0	379,654	0
嘉兴市	33	403	668,464	354,625	12,347	3,134	1,262	56.90	0.11	55,888,420	58,163	55,888,420	58,163	0	0	0	0
湖州市	29	169	309,021	11,446	223	485	9	1.82	0.08	2,131,880	378	404,213	0	0	0	1,727,667	378
海宁县	29	264	563,827	148,810	3,534	940	504	20.65	0.08	10,229,080	1,112	10,229,080	1,112	0	0	0	0
长兴县	29	252	272,998	28,348	176	866	14	2.82	0.06	19,094,609	336,634	8,232,816	62,610	0	0	10,861,793	274,024
桐乡县	34	273	509,525	183,126	4,834	3,615	642	18.99	0.14	27,440,498	11,235	27,440,498	11,235	0	0	0	0
平湖县	23	268	435,484	239,082	6,599	7,325	442	29.32	1.65	27,480,734	0	27,480,734	0	0	0	0	0
德清县	12	51	75,939	4,006	146	23	2	0.20	0.01	6,727,510	9,020	36,521	0	0	0	6,690,989	9,020
安吉县	30	147	255,548	34,855	448	471	77	3.40	0.21	25,126,810	227,738	0	0	0	0	25,126,810	227,738
嘉善县	24	329	347,716	245,774	13,181	3,661	1,094	64.66	0.42	66,520,299	14,392	66,520,299	14,392	0	0	0	0
海盐县	19	166	293,029	101,638	5,328	1,593	427	26.50	0.02	13,428,756	12,981	13,428,756	12,981	0	0	0	0
绍兴市	37	260	216,977	24,910	383	402	206	5.77	0.04	6,779,965	375	1,159,216	0	0	0	5,620,749	375
上虞县	18	104	74,453	4,002	165	24	15	1.04	0.11	586,535	230	341,374	0	0	0	245,161	230
嵊　县	52	657	440,345	61,316	601	1,034	101	5.80	0.06	32,654,462	428,528	0	0	0	0	32,654,462	428,528
新昌县	19	165	108,585	4,412	135	252	16	0.59	0.11	7,078,462	56,482	0	0	0	0	7,078,462	56,482
诸暨县	37	350	246,992	21,051	228	726	224	5.78	0.82	16,774,533	17,985	5,737,107	13,320	0	0	11,037,426	4,665
宁波市	3	3	2,391	64	2	0	0	2.54	0	23,700	0	12,230	0	0	0	11,470	0
镇海县	1	1	966	171	0	2	0	15.75	0.20	8,473	0	8,473	0	0	0	0	0
鄞　县	20	90	95,382	5,155	120	32	3	1.84	0.07	712,485	633	284,742	0	0	0	427,743	633
余姚县	20	71	66,100	3,827	63	118	4	4.04	0.18	317,371	80	275,601	80	0	0	41,770	0
慈溪县	3	4	4,052	251	0	0	0	0.43	0	3,560	0	250	0	0	0	3,310	0
奉化县	9	65	76,853	3,412	31	14	1	3.43	0.02	650,612	7,911	0	0	0	0	650,612	7,911
宁海县	17	103	63,915	3,952	56	38	3	0.72	0.06	1,676,626	6,610	0	0	0	0	1,676,626	6,610
黄岩县	12	63	45,809	5,186	65	17	8	2.20	0.04	1,470,691	866	0	0	0	0	1,470,691	866
临海县	2	4	1,484	221	19	144	0	16.64	9.70	5,917	0	1,803	0	0	0	4,114	0
天台县	17	147	93,162	10,993	508	46	29	7.16	0.15	12,180,240	10,992	0	0	0	0	12,180,240	10,992
缙云县	3	9	7,709	155	1	0	0	4.21	0	210,197	135,125	0	0	0	0	210,197	135,125

Notes: The figure is a scanned schistosomiasis prevalence statistical table of Zhejiang province from the China Schistosomiasis Atlas (Qian 1988).

C Two-Way Fixed Effects Model with Heterogeneous Treatment Effects

Equation (1) is essentially a two-way fixed effects model studied in De Chaisemartin and d'Haultfoeuille (2020). They show that if the treatment effect is heterogeneous across counties and cohorts, the two-way fixed effects estimator may be a misleading estimate of the average treatment effect even when the standard parallel trend assumption is satisfied. Therefore, we estimate a corrected difference-in-differences estimator DID_m, which allows heterogeneous treatment effects across groups and periods.

We start from a general setup. All observations can be divided into J groups and T periods. For every $(j,t) \in \{1,...,J\} \times \{1,...T\}$, let $N_{j,t}$ denote the number of observations in group j at period t. For every $(i,j,t) \in \{1,...,N_{j,t}\} \times \{1,...,J\} \times \{1,...T\}$, let $D_{i,j,t}$ and $y_{i,j,t}$ denote the treatment status and outcome of observation i in group j and period t. Then for all (j,t), let

$$D_{j,t} = \frac{1}{N_{j,t}} \sum_{i=1}^{N_{j,t}} D_{i,j,t},$$

$$Y_{j,t} = \frac{1}{N_{j,t}} \sum_{i=1}^{N_{j,t}} y_{i,j,t}.$$

$D_{j,t}$ and $Y_{j,t}$ denote the average treatment and observed outcome in group j at period t, respectively. Suppose $D_{j,t}$ can take values in $D = \{0,...,\bar{d}\}$. To define DID_m, we introduce, for all $(d,d',t) \in D^2 \times \{2,...T\}$,

$$DID_{d,d',t} = [1\{d < d'\} - 1\{d' < d\}] \begin{bmatrix} \sum_{(j,t):D_{j,t}=d',D_{j,t-1}=d,t\geq 2} \frac{N_{j,t}}{N_{d,d',t}} [Y_{j,t} - Y_{j,t-1}] \\ -\sum_{(j,t):D_{j,t}=D_{j,t-1}=d,t\geq 2} \frac{N_{j,t}}{N_{d,d',t}} [Y_{j,t} - Y_{j,t-1}] \end{bmatrix},$$

where $N_{d,d',t} = \sum_{j:D_{j,t}=d,D_{j,t-1}=d} N_{j,t}$ is the number of observations whose treatment status changes from d to d' at time t. Denote $N_{D,S} = \sum_{(j,t):t\geq 2} N_{j,t} |D_{j,t} - D_{j,t-1}|$, then

$$DID_m = \sum_{t=2}^{T} \sum_{(d,d')\in D^2, d\neq d'} \frac{N_{d,d',t}}{N_{D,S}} DID_{d,d',t}.$$

DID_m is a weighted average of a series of regular DID estimators. For a regular DID estimator $DID_{d,d',t}$, we compare the evolution of the outcomes between treatment groups and control groups in period t. The treatment level changes from d to d' in treatment groups, while the control groups keep the treatment level at d unchanged. DID_m combines this kind of regular DID estimators across all possible values of d, d', and t.

Using De Chaisemartin and d'Haultfoeuille's notation, our setup has two periods: before and after the schistosomiasis control program ($t = 1, 2$) and 858 groups ($j = 1, 2, ..., 858$). It is a sharp two-way fixed effects model, which means that all people have the same treatment variable within a group-period cell (i.e., $D_{i,j,t} = D_{j,t}$ for all $i \in \{1, ...N_{j,t}\}$). Then $D_{j,1} = 0$ for all j since no county received treatment before the start of the deworming program. $D_{j,2} = sch_j$ as the pre-control schistosomiasis infection rate is the treatment level of county j. Since there are only two periods and one initial treatment level, $N_{d,d',t} = N_{0,d',2} = \sum_{j:D_{j,2}=d'} N_{j,2}$, which is the number of observations born after 1957 in counties with $sch_j = d'$. Denote the set of endemic counties E and the set of uninfected counties U. We can write the number of people born after 1957 in uninfected counties as $N_{U,2}$. Then we can express the regular DID estimator as:

$$DID_{d,d',t} = DID_{0,d',2} = \sum_{j:sch_j=d'} \frac{N_{j,2}}{N_{0,dl',2}} \left[Y_{j,2} - Y_{j,1}\right] - \sum_{j \in U} \frac{N_{j,2}}{N_{U,2}} \left[Y_{j,2} - Y_{j,1}\right].$$

Finally, the new estimator, DID_m, can be written as:

$$DID_m = \sum_{d' \in D} \frac{N_{0,d'2}}{N_{D,S}} DID_{0,d',2}$$

where $N_{D,S} = \sum_{j \in E} N_{j,2} D_{j,2}$. The above formula clearly shows that DID_m is a weighted average of regular DID estimators comparing the evolution of the outcome in endemic counties where the pre-control infection rate is d' and in uninfected counties where the treatment is always 0, across all possible values of d'.

This new estimator relies on a normal parallel trend assumption and a stable group existence assumption. That is, if there is one treatment group, then there exists at least one control group. Our design clearly satisfies this assumption because there are more uninfected counties than infected counties. Table C1 displays the estimated results of DID_m. Standard errors are clustered at the county level and computed by 1,000 bootstrap replications. The results are similar to the cohort analysis and IV estimates. We find a larger significant increase in individuals' educational attainment compared to the baseline results. If one county with a 30 percent schistosomiasis infection rate eliminated the disease, rural women in this county were 11.61 and 16.53 percentage points more likely to complete primary school and junior high school, respectively. In terms of years of schooling, they received roughly additional one year. The positive education effect was slightly smaller for rural males, which is consistent with our results using other methods.

Table C1. Two-Way Fixed Effects Estimator with Heterogeneous Treatment Effects

	(1) Schooling (years)	(2) Literacy	(3) Primary school	(4) Junior high	(5) Employed
	Panel A. Female				
sch × post	3.584	0.036	0.387	0.551	-0.025
	(0.595)	(0.060)	(0.073)	(0.098)	(0.020)
Observations	350,149	350,149	350,149	350,149	350,149
	Panel B. Male				
sch × post	2.786	-0.049	0.260	0.445	-0.002
	(0.401)	(0.034)	(0.034)	(0.067)	(0.004)
Observations	361,523	361,523	361,523	361,523	361,523
County FEs	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports results from estimating a two-way fixed effects model with heterogeneous treatment effects, as described in De Chaisemartin and d'Haultfoeuille (2020). The dependent variables are years of schooling and four dummy variables indicating literacy, completing primary school, completing junior high school, and being employed. We add county fixed effects and cohort fixed effects and control for ethnicity in all regressions. Standard errors are clustered at the county level and computed by 1,000 bootstrap replications.