The Impacts of Hookworm Eradication in the American South. A replication study of Bleakley (*The Quarterly Journal of Economics*, 2007)

David Roodman[*]

International Journal for Re-Views in Empirical Economics, Volume 2, 2018-3, DOI: 10.18718/81781.7

JEL: I18, O15
Keywords: worms, public health and economic development, replication study

Data Availability: This paper uses data sets from the IPUMS project, all of which are easily obtainable from usa.ipums.org. Some of these data sets—the 100% U.S. census samples for 1910 to 1940—are not licensed for public redistribution. The redstributable data sets, Stata do files and log files, and detailed instructions for the access to the all data are available from the website of the journal www.iree.eu.

Please Cite As: David Roodman (2018). The Impacts of Hookworm Eradication in the American South. A replication study of Bleakley (The Quarterly Journal of Economics, 2007). *International Journal for Re-Views in Empirical Economics*, Vol 2(2018-3). DOI: 10.18718/81781.7

Abstract

Through designs akin to difference-in-differences, Bleakley (2007) produces evidence that the campaign to eradicate hookworm from the American South circa 1910 boosted school enrollment in childhood and income in adulthood. This comment works to replicate and reanalyze that study. Innovations include incorporation of the larger U.S. Census samples now available, and fitting of specifications focusing more sharply on the timing of any effects of the campaign, which are the basis of the most credible identification. The long-term convergence between historically low- and high-hookworm areas documented in Bleakley (2007) began decades before the campaign and did not accelerate in a way that would invite hookworm eradication as an explanation. Likewise, in the case of adult income, the convergence continued for decades after. In sum, hookworm eradication did not leave a telltale imprint on the historical record assembled here.

[*]Senior Advisor, GiveWell, 182 Howard St., #208, San Francisco, CA; Senior Advisor, Open Philanthropy Project, 182 Howard St., #225, San Francisco, CA. david.roodman@givewell.org. GiveWell and the Open Philanthropy Project have no interests in the topic of this paper that would constitute a conflict of interest. I thank Hoyt Bleakley, Caitlin McGugan, and three anonymous reviewers for comments on earlier drafts; Christian Smith and Claire Wang for help with data collection; and Zachary Tausanovitch for help with data collection, database construction, and analysis.

Received January 22, 2018; Revised August 24, 2018; Accepted October 5, 2018; Published December 10, 2018.

1 Introduction

Deworming children in developing countries is cheap. The medicines—albendazole for geohelminths, praziquantel for water-transmitted schistosomiasis—are practically free. Administering the drugs is inexpensive, if done where children already gather, such as at schools. As a result, the charity evaluator GiveWell estimates that mass childhood deworming costs $0.32 per dose in India and $0.79 in Kenya.[1]

The *benefits* of mass deworming are harder to gauge. The evidence base on the short-term impacts on nutrition and cognition is rich enough to have supported several meta-analyses (Taylor-Robinson et al. 2015; Welch et al. 2016; Croke et al. 2016). These have disagreed in their assessment: the first two cited are pessimistic about any overall benefit but the third is more optimistic, with regard to the well-studied outcome of child weight. *Long-term* impacts could cumulatively matter far more; but here the evidence is thinner. In an approximately ten-year follow-up on the influential Miguel and Kremer (2004) deworming experiment in western Kenya, Baird et al. (2016) find impacts on earnings sufficient to generate an internal rate of return of at least 32% per annum. Ozier (2018) follows up on the same experiment at about the same time, and reports cognitive gains among children who were too young to have directly participated in the experiment but who could have benefited indirectly, through the deworming of their school-age siblings and neighbors. Croke (2014) examines impacts on academic outcomes in a ten-year follow-up on a randomized deworming trial in Uganda.

For decisionmakers trying to assess the effects of deworming, the paucity of modern, experimental evidence on the long-term consequences raises the importance of one noted historical study. Bleakley (2007) evaluates the Rockefeller Sanitary Commission's campaign to eradicate hookworm from the American South circa 1911–14. Through designs akin to difference-in-differences (DID), the study identifies impacts off of the interaction of two sources of variation: geographic differences in the initial prevalence of hookworm, and the particular timing of the campaign. The Bleakley (2007) results parallel those from Kenya (Miguel and Kremer 2004; Baird et al. 2016): mass deworming of children boosted schooling in the short run and earnings in the long run.

Bleakley (2007) has been influential. It had accrued 749 citations in Google Scholar as of August 21, 2018. Among papers that pertain to health, education, and welfare and had appeared between 1991 and 2014 in the top-five economics journals, Bleakley (2007) ranks among the 50 with the highest citation rates (Linnemer and Visser 2017). GiveWell used to cite it, along with Baird et al. (2016) and Croke (2014), as key evidence in favor of donating to deworming charities.[2]

The present paper replicates and reanalyzes Bleakley (2007). It returns to primary sources, constructs new data sets modeled on the originals, and strives to reproduce nearly all the original tables and figures.[3] Moving from replication to reanalysis, the paper then modifies specifications in order to test robustness and zero in on questions that are relevant for assessing any attribution of

[1] GiveWell, "Deworm the World Initiative, Led by Evidence Action," November 2016, `web.archive.org/web/20170918025142/http:/www.givewell.org/charities/deworm-world-initiative`.

[2] See GiveWell, "Combination deworming (mass drug administration targeting both schistosomiasis and soil-transmitted helminths)," archived March 7, 2016, at `web.archive.org/web/20160307215042/http:/givewell.org/international/technical/programs/deworming`.

[3] The exception is Figure I, which is preliminary to the main analysis and mostly uses separate data.

impacts. In particular, the paper:

1. *Takes advantage of the larger census samples now available from the Integrated Public Use Microdata Series* (IPUMS; Ruggles et al. 2015). This expansion includes 100% "samples" for 1910–40.

2. *Copies specification choices among the displays to test robustness.* For example, where, in the original, a table tests for impacts on three outcomes and the corresponding figure illustrates for only one, analogous figures are here generated for all three.

3. *Performs tests to focus more sharply on whether trends break at times explicable by the eradication campaign*; for the timing of the eradication campaign—its historically contingent start date and rapid execution—is the most credibly exogenous source of identifying variation in the study.

4. *Addresses a few econometric issues.* For example, the three-stage estimation process in one analysis is revised to factor uncertainty from initial stages into the final stage's standard errors.

A pre-analysis plan, registered with the Center for Open Science, envisions some of these steps—the fourth and part of the third. As that statement implies, I did not confine this study to steps articulated in the pre-analysis plan. But the plan does make transparent which steps I chose before encountering the data.

The new analysis focuses on figures more than tables. The figures bring out temporal patterns clearly and motivate formal tests mentioned above in item 3.

The replication attempt reported here recognizably matches nearly all the original's tabulated and graphed results. The largest exception is that a prominent graphical finding in the original—a striking school enrollment jump between 1910 and 1920 in historically high-hookworm areas relative to low-hookworm areas—is now harder to distinguish from the pre-treatment trend.

Moving from replication to reanalysis, I present new results that further question the Bleakley (2007) conclusion that hookworm eradication brought detectable short- and long-term benefits. As a first step, I replicate the school enrollment graph, just mentioned, for the other Bleakley (2007) indicators of human capital investment, namely, full-time school attendance and literacy. And I test for robustness to using newer, larger census microdata samples. Expanding the census samples produces results that more strongly suggest that the relative upward trend in schooling and literacy in historically high-burden areas began as early as 1870. And a long-term trend beginning well before the eradication campaign is hard to attribute to the campaign.

The findings are similar for long-term impacts on income. The replication's birth-cohort-by-birth-cohort results confirm that income—more precisely, the "occupational standing" variables used to proxy for it—gradually converged across the gradient of baseline hookworm burden. But when I formally test whether convergence temporarily accelerated at a time associated with the eradication campaign, I do not find convincing evidence in favor. Again, the convergence begins decades earlier than would be expected from hookworm eradication efforts. And it continues later.

I began corresponding with Hoyt Bleakley about these findings in May 2017. Bleakley stated that the original data and code are effectively lost. In January 2018, I received substantive comments from Bleakley on an earlier version of this paper, which are incorporated here.[4]

Section 2 of this paper details the Bleakley (2007) designs. Section 3 introduces several cross-cutting themes in the reanalysis. Sections 4 and 5 replicate and reanalyze the short-term and long-term impact regressions. Section 6 concludes.

2 The Bleakley (2007) designs

To study the campaign to eradicate hookworm disease from the American South in the early 1910s, the Bleakley (2007) specifications combine three sorts of variables:

- Cross-sectional variables, observed once per geographic unit. These include indicators of pre-eradication hookworm prevalence (H), along with controls relating to health, education, race, agriculture, and parental background. All come from sources published about a century ago.
- A time series indicator for exposure to the post-eradication regime (Exp). This takes the same values in all geographies, regardless of baseline hookworm burden. It is interacted with H to form the treatment proxy.
- Variables built from decennial census microdata (Ruggles et al. 2015). These include demographic controls—age, sex, race—and the outcome measures such as school enrollment and occupational standing. These vary geographically and temporally.

In one case, discussed below, the data are aggregated before analysis, within birth state–birth year–census year cells. The rest of the regressions are run directly on census microdata.

For all Bleakley (2007) designs, the estimating equation can be written

$$Y_{ijt} = (H_j \times Exp_t)\beta + \mathbf{z}'_{ijt}\boldsymbol{\alpha} + \mathbf{x}'_{jt}\boldsymbol{\gamma} + \delta_j + \delta_t + \epsilon_{ijt} \tag{1}$$

for outcome Y for individual i in geographic unit j at time t. β is the impact parameter of interest. The δ_j and δ_t are place and time dummies, and obviate the inclusion of H_j and Exp_t as controls. The $\mathbf{z}_{ijt}$ are individual-level demographic traits such as age, sex, race, and interactions thereof.

[4] Since the November 21, 2017, version of this paper, these changes have been made: fixing a bug substantially reducing 1940–50 samples in SCS estimates including those years; dropping year-2000 census data from multicensus RC regressions because of a quality concern raised by Bleakley, thus conforming with Bleakley (2007) while departing from Bleakley (2010); implementing an exclusion—of women—from the multicensus RC samples, documentation of which was lost in the editing of Bleakley (2007); quintupling weights in 1950 census data for SCS education regressions to compensate for schooling only being observed in 20% of cases; correcting (making uniform) the weights in the replication of the original's 0.4% 1910 census sample, which in the modern IPUMS interface is extracted from a larger (1.4%) sample with a more complex weighting scheme; in SCS data sets, replacing log change in health spending with change in health spending, imputing zero for missing observations of number of patients examined and number treated, adding (documented) controls for 1910 urban fraction and dummies for missingness in parents' occupational incomes scores, obtaining data for log change in value of school plant and equipment in Mississippi, and imitating the (previously overlooked) imputation for school term lengths in Kentucky in the pre-treatment period.

The $\mathbf{x}_{jt}$ are not true panel variables, in the sense of being observed in primary sources in multiple times and places. Rather, all are products of pure cross-sectional and pure time series variables. An example is the set of interaction terms $\delta_j \times t$, which is included in some specifications to control for area-specific linear time trends.

Bleakley (2007) also performs graphical analyses, which involve running a version of (1) separately for each t-indexed cross-section:

$$Y_{ijt} = H_j\beta_t + \mathbf{z}'_{ijt}\boldsymbol{\alpha}_t + \mathbf{x}'_j\boldsymbol{\gamma}_t + \epsilon_{ijt} \tag{2}$$

The $\mathbf{x}_j$ are optional cross-sectional controls. These regressions yield a series of coefficients, β_t, which measure the (conditional) cross-sectional association between baseline hookworm prevalence and the outcomes. Bleakley (2007) then conducts inference about whether the β_t series constitutes evidence of impact, in the sense that Exp is a good explanator for the series. The set of regressions (2) can also be performed as a single, full-sample regression in which time dummies δ_t are interacted with all the right-side variables.

Bleakley (2007)'s two designs, successive cross-section (SCS) and retrospective cohort (RC), differ in how they group the data—in effect, in what they take the indexes j and t to refer to. These choices in turn shape the definitions of H and Exp (baseline prevalence and potential campaign exposure).

The SCS design categorizes an observation by when and where it was collected, meaning the census year and the person's place of residence. Exp is then simply a dummy for post-campaign censuses, i.e., for $t \geq 1920$. As for H, the census observes place of residence with high precision—though the public, digitized microdata less so. In principle, this allows the SCS specifications to take full advantage of the county-level spatial resolution in the Rockefeller Sanitary Commission's (RSC's) baseline hookworm prevalence surveys.[5] H could be defined by county of residence and linked to other variables at that level. In practice, Bleakley aggregates baseline prevalence and the other county-level variables to the "state economic area" (SEA; Bogue 1951). Each SEA consists of several contiguous counties within a state. SEAs are attractive because they are more stable than counties, which have sometimes merged or split or had boundaries redrawn. Also, starting in 1950, IPUMS census records report residence by SEA but not county. Thus, in the SCS design, the geographic index j refers to SEAs. Since the RSC waged its campaign across 11 southern states from Virginia to Texas, it surveyed prevalence only in those states. This restricts the SCS regressions to SEAs in those 11 states.

In the retrospective cohort (RC) design, t and j index time and place of birth instead of time and place of survey. This structure facilitates assessment of long-term effects by minimizing attrition from migration. If a person was born in Georgia in 1915 just after the eradication campaign, and showed up in successive censuses as a bricklayer in Atlanta in 1940, a general contractor in Lexington in 1950, and a construction company director in Phoenix in 1960, all three observations would be associated with Georgia, 1915. Bleakley (2007)'s single-census RC specifications use data from the 1920 or 1940 census. The multicensus specifications pool all census data from 1870 to 1990 that were available to Bleakley.

[5] In fact, the RSC subdivides a few counties for reporting purposes.

The redefinition of the time and place indexes in moving from the SCS to the RC design triggers several other changes. Partly because the cadence of t shortens from decadal to annual, Bleakley (2007) incorporates more timing information into Exp. Instead of being a post-campaign dummy, Exp now measures the number of childhood years of exposure to the post-campaign regime. For this purpose, the campaign is taken to begin in 1910 and childhood to end at age 19. "Nineteen is chosen because most individuals in this period would have completed their schooling by that age, and hookworm infection was negligible at older ages" (Bleakley 2007, p. 95). Thus, Exp=0 for a person born in 1891, who would have reached 19 in 1910 and just missed the opportunity to enjoy the post-eradication regime during childhood. Exp=1 for those born in 1892, 2 for those born in 1893, and so on. It is 19 for all people born in or after 1910. I will call this piecewise-linear function of birth year the step function. It implies the assumption that exposure at each year of childhood matters equally for long-term outcomes.

Meanwhile, since the census observes place of birth only at the state level, in the RC regressions, the state replaces the SEA as the geographic unit. As a result, to perform the RC regressions, Bleakley (2007) widens the geographic scope to the continental United States. And the study drops the county-level Rockefeller prevalence surveys as the basis for H in favor of a state-level indicator with national coverage (Kofoid and Tucker 1921).

3 Themes in the replication and reanalysis

3.1 Pre-analysis plan

A pre-analysis plan for this replication and reanalysis was registered with the Center for Open Science (osf.io/yb537). As noted, it does not confine the analysis. But it discloses which parts were pre-conceived and which were chosen after encountering the data. Here are the steps envisioned in the plan, along with commentary:

- "Testing for sensitivity to any data or coding errors exposed in the original." None were exposed, for lack of access to the original data and code.
- "Performing two-stage least squares instead of the original's indirect least squares [ILS] in order to obtain proper confidence intervals for instrumental variables point estimates." This step was ill-conceived. The original uses ILS where conventional instrumental variables estimation is impractical, because, e.g., the impacts of the instruments on the treatment and on the outcome are estimated in different contexts.
- Performing "pure time-series versions of the sequential cross-sections (SCS) analysis, in which samples are restricted to areas of above-average baseline prevalence." This was done (see section 4.2 below). Since the (temporal) variation in Exp is more credibly exogenous than the variation in the other component of treatment, H, a pure time series specification seemed worthwhile as a robustness check.
- "More-conservative error-clustering choices, such as clustering county-level estimates by state rather than State Economic Area." (By "county-level estimates," the SEA-level SCS regressions are meant.) For the SCS regressions, clustering was not expanded from SEA to state, because it seems rather demanding when the sample has only 11 states; also, even with SEA-clustered standard errors, the reanalysis casts substantial doubt on the original. However, the

reanalysis of the multicensus RC regressions does move from clustering by birth state–birth year combination to clustering by state, across time, in order to address serial correlation.

- "Re-doing the two-stage assessment of whether the hookworm campaign helps explain the convergence in long-term earnings between low-and high-prevalence areas (equation 5 and Table VI) in a way that factors the uncertainty of the estimates from the first stage into the second, either analytically or by bootstrapping." This was done, and is reported in section 5.2. In fact, the "two-stage assessment" has three stages, which that section also explains. The alternative adopted here is to combine all stages into a single ordinary least squares (OLS) regression on microdata, as in the rest of Bleakley (2007).

3.2 Expanded IPUMS samples

The coverage of the IPUMS U.S. census data archive has expanded steadily over the years, both in the rounds included and in the size, or "density," of samples digitized. Bleakley (2007) reports last obtaining IPUMS data on May 30, 2003, for the SCS analysis; on February 5, 2003, for the single-census RC; and on November 14, 2005 for the multicensus RC. Bleakley (2007) largely does not specify the densities of the samples used, but they can be estimated by reviewing the history of https://ipums.org/usa/sampdesc.html at archive.org, as well as the change log at usa.ipums.org/usa-action/revisions. Table 1, column 1, shows my estimates. Encouragingly, certain sample sizes reported below for the base SCS regressions nearly match corresponding values reported in an early version of Bleakley (2007).[6]

In addition to reconstructing the original data set according to these estimates, I test robustness by switching to newer, larger IPUMS samples. (See column 2 of Table 1.) The "expanded" collection of samples adds microdata from 1860 and 1930—though the 1860 data figure only in the multicensus RC regressions.[7] Density rises to 5% in 1900 and 1960, and to 100% for 1910–40, using preliminary releases for the latter.[8] However, although the census asked respondents about their literacy through 1920, the preliminary 100% samples for 1910–20 lack this information; so for SCS literacy regressions, the next-largest IPUMS samples are used, namely, the 1.4% for 1910 and 1% for 1920.

While the data expansion was not pre-registered, it was to a degree inevitable since Bleakley (2007) does not fully document the original samples used and because the modern IPUMS interface tends to hide two samples that Bleakley (2007) appears to use: the 1-in-760 sample for 1900 and

[6] In the Bleakley (2002) job market paper, Tables III and VI report observations counts. As shown below in Table 4, panel A, the replication matches these well. Similarly for the 1910–20 adult literacy SCS regressions (Table IV in the original, Table 6 here).

[7] A previous version of the present paper added 2000 data to the multicensus RC regressions as well, like the Bleakley (2010) malaria study. However, in January 2018, Hoyt Bleakley pointed me to a "user caution" on the IPUMS website about the occupation codings in the 2000 census, which are the basis for the occupational income score and Duncan's socioeconomic index (usa.ipums.org/usa-action/variables/OCCSCORE). I have therefore dropped the 2000 data. This hardly affects results.

[8] IPUMS staff state that the 100% samples are "preliminary" in the sense of being incomplete, not unreliable. Some variables are still being added. See answers.popdata.org/How-interpret-preliminary-label-full-count-data-q2390259.aspx

1-in-250 sample for 1910.[9] For concision, I only perform the data expansion robustness test on the Bleakley (2007) figures, not the tables. (Section 3.4 explains the focus on the figures.)

All new regressions reported below incorporate person-level sampling weights provided by IPUMS. Most IPUMS samples are "flat," meaning that they statistically represent the population without weighting. But the 1950 file over-samples large households, and the 1990 5% file used here over-samples small communities (Ruggles et al. 2015; usa.ipums.org/usa/intro.shtml#weights), both of which selection traits could be endogenous to occupational standing, an outcome central to Bleakley (2007).[10] And since different censuses are sampled at different densities, pooling them unweighted would introduce imbalances. Bleakley (2007) does not mention using sampling weights.

3.3 Differences among specifications

Nearly all the Bleakley (2007) results appear in tables and figures. Naturally, the specifications behind these displays vary in certain respects. The differences generate some minimally arbitrary robustness tests: distinctive choices in one display can be copied to others.

The Bleakley (2007) figures and tables are listed here in Table 2. The internal differences revealed by perusing that table are of two sorts. Some are in a sense presentational. For example, using SCS regressions, Bleakley (2007)'s Tables II and III examine impacts on school enrollment, full-time school attendance, and literacy, with and without various control sets. The parallel Figure II looks only at one outcome, enrollment. And it does so without including any extra controls. In contrast, the corresponding RC figure, Figure III, depicts runs with and without controls. Generating the "missing figures" implied by these contrasts—for all three SCS outcomes, with and without full controls—does not appear to raise major substantive issues, and is done below.

A deeper fissure separates Bleakley (2007) multicensus RC results—in Figure III and Table VI—from the rest of the paper. For these displays, the underlying analysis is run on aggregate data rather than microdata. And, more significantly, the samples are restricted to white men, whereas all the other Bleakley (2007) analysis include blacks and women.[11] Whether these restrictions best serve estimation and inference is not merely a question of presentation. The next section discusses the issue more. The expanded-data results reported below do add blacks and women to the multicensus RC regressions, while graphs in the appendix examine robustness of the new findings to not adding them.

[9] The Preston 1-in-760 sample is available at usa.ipums.org/usa/samples.shtml. The old 1910 sample is marked within the newer 1.4% sample by the field SAMP1910.

[10] School enrollment was a "sample-line" variable in 1950, meaning that it was only collected for a fifth of individuals. For these individuals, the sampling procedure is in fact flat, with a density of 0.2%. Schooling regressions reported here use the IPUMS variable SELWT for 1950 data instead of the usual individual-level weighting variable PERWT.

[11] These exclusions are easy to miss. The main text does not mention neither. The data appendices mention the restriction to whites in one place. In January 2018, Hoyt Bleakley pointed me to the working paper version, Bleakley (2006), which has little search engine presence but does document both exclusions.

3.4 On identification in Bleakley (2007)

The Bleakley (2007) treatment proxy, $H_j \times Exp_t$, is the product of two factors, one geographic, one temporal. The geographic factor, initial hookworm infection prevalence, is not credibly exogenous. Within the South—the sampling region for the SCS regressions—low- and high-hookworm areas differed systematically. Bleakley writes:

> Hookworm larvae were better equipped to survive in areas with sandy soil and a warm climate. Broadly, this meant that the residents of the coastal plain of the South were much more vulnerable to infection than were those from the piedmont or mountain regions. (p. 79)

As a result, low- and high-hookworm areas may have differed in other respects too—in crops historically grown; in suitability for the peculiar institution of slavery; in wealth, inequality, and education. It would also not be surprising if the economic importance of these historical differences dwindled over the long study period embraced here, as agriculture's share in the economy shrank.

That brings us to the second factor in $H_j \times Exp_t$, which is temporal. As a source of identification, Exp can be more credibly exogenous than H, but more so at the scale of the decade than the century. That hookworm was beaten back between, say, 1850 and 1950 is a marker for the scientific revolution occurring during this time, and thus for many developments not fully controlled for in the Bleakley (2007) regressions. That sharp progress occurred against the disease in the early 1910s rather than five years before or after is more an accident of history. This is why the most compelling evidence that can emerge from the Bleakley (2007) analysis would be not merely of long-term convergence in schooling or future occupational standing. Rather, in the spirit of regression discontinuity and regression kink designs, the most compelling evidence would be of acceleration and deceleration in such convergence coinciding with the eradication campaign.

This observation—of the primacy of short-term effects—generates several implications for the reanalysis carried out here.

First, it tends to challenge the Bleakley (2007) exclusion of blacks and women from the multicensus RC regressions, the ones that estimate impacts on career-averaged occupational standing. Bleakley (2007) does not motivate the exclusions. However, Bleakley (2010) does explain the restriction to whites in nearly the same analytical context, studying malaria rather than hookworm eradication:

> I focus on US whites for several reasons. First, only a small proportion of blacks lived outside of the most malarious states among the earlier cohorts, which means that they make for an imprecisely measured point of comparison. Second and more importantly, that same population of blacks was less likely to have been enslaved, which means that they make for an inappropriate control group for those blacks born into slavery in the malarious south. The estimates reported below (for whites) are similar to those obtained if I include native blacks in the base sample. Estimates using blacks only, however, are imprecise and sensitive to control sets employed.

In a corrigenda, Bleakley (2018) clarifies that women were excluded for similar reasons. The role of both these large demographics in the labor market changed radically in the 1870–1990 span of the multicensus RC data sets. Bleakley (2018) elaborates:

> ... there is a tradeoff here in the long-term analysis. Those people who were born only a decade or two apart are more likely to be comparable to each other, but unlikely to be useful in sorting out the cross-cohort timing of income convergence. I made the judgment call that this comparability problem was too severe in the case of blacks because of enslavement at the outset of the sample, their distinct regional distribution over time, and later the effects of the increasing integration of blacks into the mainstream economy.

But if evidence of shorter-term changes is most potentially compelling, as argued here, then inference is better served by making this tradeoff oppositely. The preferred specifications presented here include blacks and women in order to increase statistical power to detect shorter-term changes. (In practice, just as reported in the Bleakley (2010) quote above, this demographic expansion does not materially affect results.)

The second implication of the interest in shorter-term changes is to amplify a broader concern about the Bleakley strategy of regressing century-scale time series on Exp, an indicator whose activity is contained within a 19-year spell: it could generate spurious results. As an example, if the century-scale trend in the schooling gap between historically low- and high-hookworm areas is S-shaped, with the gap narrowing over many decades, then regressing it on the Exp step function could falsely assign explanatory power to the term. Controlling for time linearly or quadratically may not suffice to remove the ambient S-shaped trend, since the parabola is a poor model for the logistic curve. Increasing long-range comparability of samples by restricting to white men mitigates this concern by reducing potential confounding. But, as already suggested, it does not eliminate the concern because even among white men it is plausible that long-term trends operated that are not removable by quadratic time controls. This worry thus pertains to all the tabulated regressions in Bleakley (2007), since all control for time at most quadratically. The expansion carried out here, to blacks and women, exacerbates the issue since their labor market experiences substantially evolved for reasons separate from hookworm.

The third implication of the interest in shorter-term changes is a constructive response to the concern just aired. It is an emphasis on methods that test for the "fingerprint" of the eradication in the time dimension, often with graphical presentation. In reanalyzing Bleakley (2007), I search for the fingerprint in three ways:

- One is to regress on the Exp step function while introducing controls for time polynomials up to order five. This approach is implicitly pre-registered in that Bleakley also employs it, albeit only detailing results up to order two.[12]
- The second way, not pre-registered, does not add controls for time trends. Instead, it generalizes the Exp step function to a piecewise-linear spline. Kinks are allowed only at the times set, or most naturally implied, by Bleakley: 1910 and 1920 for the SCS regressions, which have data only for census years; and 1891 and 1910 for the RC, those two dates bracketing the period of rise in Exp. After fitting these models, I test whether the slope changes at each kink, as it should under a step-like impact contour.
- The third way is a variant of the second, and is also not pre-registered. It seeks to compensate for a limitation of the second, namely, the rigid and somewhat arbitrary pre-specification of

[12] Bleakley (2007) presents results for specifications up to order two, but footnote 25 reports testing up to order five.

the two kink dates. In analogy with the Bai and Perron (1998) approach to searching for structural breaks, it allows the two trend break dates in the linear spline model to vary along with the other parameters, and simultaneously chooses all according to the least-squares criterion. This method has the advantage of letting the data speak to where the structural breaks occur. Its weakness is that is drawn to the most important breaks, regardless of whether they occur when hookworm eradication could plausibly explain them. It is applied only to the RC regressions, in which the data are finely enough resolved in the time dimension—annual rather than decadal—to support a meaningful best-fit search for kink dates.

None of these approaches is obviously optimal, because all pre-specify a family of models in the face of ignorance about the true model. But as a group, the approaches seem intuitive and informative. And all embody the idea that the most persuasive evidence relates to timing.

4 Replication and reanalysis: Successive cross-section specifications

Recall that the successive cross-section (SCS) design groups observations by census round and place of residence. Place of residence is resolved to the state economic area, which is a cluster of counties. Coverage is confined to 11 states in the American South. The exposure variable Exp is a dummy for censuses fielded after the eradication campaign of the early 1910s.

4.1 Replicating Bleakley (2007) Tables I–IV: Short-term impacts on children and adults

Table 3, below, follows the format of Bleakley (2007) Table I in order to compare the original and reconstructed data sets on first and second moments. The table contains three pairs of columns, the first for the whole sample, and the second and third for low- and high-prevalence subsamples, demarcated by a child infection rate of 40%. Within each pair of columns, the first is copied from Bleakley (2007) Table I while the second is computed from the reconstructed data set.

Overall, the original and new SCS data sets appear to match well. For the whole sample, the mean and standard deviation of the baseline infection rate match almost exactly, as do census-sourced variables such as school enrollment and population black. The match is poorer for the variable "Individuals treated at least once," but this variable is not used in the analysis. The last four variables, relating to education, were hardest to reconstruct, because they come from periodic state government reports with diverse formats and reporting conventions, and because Bleakley (2007) does not precisely document which editions are used. So it is not surprising that the matches for them are also less precise—even if still broadly reassuring. The reconstructed data set includes two more SEAs, both of which fall into the below-40% subsample. All the discrepancies are hard to explain without access to the original data and code.

In similar fashion, Table 4 replicates the first set of Bleakley (2007) SCS estimates, from the original's Table II. Each cell reports the coefficient on $H_j \times Exp_t$ in a distinct regression. The first column pair in panel A also compares on sample size, taking the original's observation counts from the Bleakley (2002) job market paper.

For the 1910–20 difference-in-differences specification, in the first row of Table 4, the replication again matches well. Here, with reference to (1) above, the individual-level demographic traits

$\mathbf{z}$ are all interactions between, on the one hand, sex and race dummies and, on the other, Exp and a continuous age variable. Time and SEA dummies are included, but the additional control set, $\mathbf{x}$, is empty. The next two rows, the remainder of Panel A, also present reasonable matches. Of these, the first expands the sample to 1900–50. Literacy is dropped as an outcome because it is not available in the census after 1920. The next row further controls for SEA-specific linear time trends.

In panel B of Table 4, the matches start to degrade as the specifications become more demanding. All these specifications retain the SEA-specific time trends, except in the literacy regressions, where lack of data for 1930–50 reduces the number of time periods. The first row of the panel introduces state$\times$post-treatment fixed effects to control for state-level policy shocks. The next row goes further to mitigate mean reversion, controlling for the interactions between state$\times Exp$ and the 1910 state-averaged value of the dependent variable. The last row of Table 4 makes the most radical changes to the specification, and yields the poorest matches—for reasons that are, again, hard to determine. As in the RC regressions, the sample expands to the entire country and baseline infection is observed in the state of birth rather than the SEA of residence. The mean reversion control is retained. The new estimates are effectively indistinguishable from 0.

Bleakley (2007) Table III tests the results in Table II for robustness to additional controls, and explores demographic heterogeneity. To further assess and document the quality of match between original and replication data sets, these regressions are reconstructed and reported in Table 5 below. Most of the new results broadly corroborate the originals. The greatest difference appears in panel B, which adds more controls. Including the full set of controls for health, education, race, agriculture, and parental background erases the suggestion of impact on school enrollment and literacy (last row of panel B), but not on full-time school attendance.

Last among the SCS tables, Table 6 replicates Bleakley (2007) Table IV, which checks for impacts on adult outcomes: literacy, labor-force participation, occupational income score (OIS), and urban residence. (OIS is a proxy for income: it is the median income, in 1950, for a person's reported occupation, expressed in hundreds of 1950 dollars.) The design is again difference-in-differences, using 1910 and 1920 census data. Bleakley (2007) finds no robust suggestion of impacts on any of these outcomes. This negative finding fits the hookworm impact theory in that adults were infected much less than children, and stood to benefit much less from eradication. By the same token, it weakens competing explanations for the original's findings of impacts on children. If factors such as migration and income changes somehow caused the apparent impacts on children, they would tend to affect adults as well. The replication results largely line up with the original ones, except that now there is an association with gains in adult literacy (first row of Table 6).

The replication of SCS regressions provides some assurance that the old and new regressions are similar in data and method. Some discrepancies are to be expected since IPUMS is continually improving its existing census data sets and since the original Bleakley (2007) data and code are inaccessible. By and large, the original results are recognizable in the reconstructions. While the new data set, like the original, presumably contains some undetected errors, it has the virtue of being publicly accessible.

4.2 Replicating and reanalyzing Figure II: Short-term impacts on children

Having replicated the tabulated SCS regressions to assess the match between old and new, we turn next to graphical analysis of the same data. Bleakley (2007) includes a single plot based on SCS regressions. It corresponds to the upper left result of Table II in the original and Table 4 here, and is derived by separately fitting (2) to the data from each census round in 1900–50. The dependent variable is school enrollment. Only demographic controls are included. As foreshadowed, I work to replicate that graph and then introduce four innovations:

- Rendering it for the other SCS outcomes, in analogy with Bleakley (2007) Figure III.
- Also rendering it while including the full control set, again in analogy with Bleakley (2007) Figure III.
- Using the larger IPUMS samples.
- Applying several tests for the presence in the time series of a step contour of the form postulated in Bleakley (2007).

The attempted replication of Bleakley (2007)'s Figure II appears in the upper-left pane of Figure 1, below. The blue dots are point estimates and the vertical grey bars their 95% confidence intervals. Shading within the bars indicates gradations in confidence.[13] Standard errors are clustered by the census year-SEA combination.[14] Consistent with Bleakley (2007) Figure II, the cross-sectional association between baseline hookworm burden and school enrollment rises between 1910 and 1920, and more rapidly than in the periods on either side. However, while the deceleration in 1920 is sharp and statistically significant, the acceleration in 1910 is less certain. The null hypothesis that the 1900–10 slope equals the 1910–20 slope is rejected by a two-tailed t-test only at $p = 0.36$, while the analogous test around 1920 returns 0.04. In the plot, red line segments are drawn to indicate the focus on these two tests and the p-values just quoted appear at the bottom. In sum, in the replication, it is not clear that the schooling catch-up in historically hookworm-burdened parts of the South accelerated with eradication.

The rest of the first row of Figure 1 moves to the other Bleakley (2007) human capital outcomes. For full-time school attendance, both slope changes are statistically significant, and in the predicted directions. On the other hand, the 1910–20 literacy trend shows no break with the past. (Lack of data prevents checking for a literacy deceleration around 1920.)

The second row of Figure 1 adds Bleakley (2007)'s full control sets. As in Table 5, panel B, row 3, this destroys most suggestion of an impact on schooling and literacy. The sharp trend reversal in 1910 for full-time school attendance may explain why the corresponding regression in that table row returns a statistically significant coefficient.

Next, I update Figure 1 with the expanded IPUMS samples, yielding Figure 2. The graphs now tell a more consistent story: the association between historical hookworm prevalence and schooling did rise between 1910 and 1920, and rose less in 1920–30—indeed, fell. The decelerations

[13] The Bleakley (2018) corrigenda confirms that the confidence level in the original is also 95%.

[14] The "demographic controls" referred to in the caption of Bleakley (2007) Figure II are taken to be those listed in the caption for Table II: "age, female, female$\times$age, black, and black$\times$age."

around 1920 are generally statistically significant ($p = 0.10, 0.17, 0.06, 0.07$ for school enrollment with and without full controls and full-time school attendance with and without full controls). But the rises in 1910–20 did not break from the preceding decade ($p = 0.87, 0.94, 0.43, 0.68$ for the schooling outcomes just mentioned; $p = 0.88$ for literacy without controls and $p = 0.17$ for a bend the "wrong" way for literacy with controls).

In the Appendix, I treat in the same way the contemporaneous adult outcomes studied in Bleakley (2007) Table IV, and replicated above in Table 6. (See Figure A1.) The results corroborate Bleakley (2007)'s difference-in-differences findings of little apparent impact. Although the replications in Table 6 differed in suggesting an impact on adult literacy, the graphs suggest that this gain continued a pre-existing trend (right end of Figure A1).

Last, I perform time series variants of these regressions. These add little insight, but are included since they were pre-registered. I split the samples in two, by whether an SEA's baseline prevalence exceeded 40%—just as in Bleakley (2007) Table I. Within these low- and high-prevalence subsamples, I fit:

$$Y_{ijt} = \beta_t + \mathbf{z}'_{ijt}\boldsymbol{\alpha}_t + \mathbf{x}'_j\boldsymbol{\gamma}_t + \epsilon_{ijt} \tag{3}$$

$\mathbf{x}$ is empty in the basic specification and holds all the cross-sectional variables in the full-controls specification. These specifications are motivated by the idea that pure time series evidence of sharp, appropriately timed gains in schooling and literacy would strengthen the attribution to the eradication campaign. In fact, we find similar trends in both groups. (See Figure A2, which uses the expanded data set for precision.)

Overall, the replication and extension of Figure II substantially weaken the case that hookworm eradication boosted human capital investment in children. The relative gains in historically hookworm-burdened areas during 1910–20 appear to have continued pre-existing trends.

5 *Retrospective cohort specifications*

As explained in section 2, the RC specifications group census observations by state and year of birth. The geographic coverage expands from the South to the continental United States. The exposure variable Exp now takes the step function form with respect to birth year, holding flat at 0 through 1891, rising linearly through 1910, then flattening again. Each tabulated RC regression in Bleakley (2007) takes data from a single census. The corresponding figure fits to data from many censuses at once.

Since the controls are all observed at the state level, the primary sources are more consistent and complete than some of the sources of county-level information for the SCS regressions. Still, ambiguities surface here too, which again impede exact reconstruction of the analysis data set. Several education variables take data from federal reports for "circa 1902–32," so the original and reconstructed data sets may use different editions. I could not see how to construct one control, male employment in 1930, from the cited source, ICPSR (1984), so I turned to the primary source, as instantiated in the 1930 IPUMS 100% census sample.[15]

[15] ICPSR (1984) offers the unemployment denominator V131, "number of gainful workers" in 1930, but this is not subdivided by sex.

5.1 Replicating Table V: Single-census retrospective cohorts

Bleakley (2007)'s single table of RC regressions, Table V, is replicated below in Table 7. It assesses impacts on three outcomes: log earnings and years of schooling as reported in the 1940 census; and adult literacy as reported in the 1920 census. Earnings and schooling regressions are restricted to ages 25–60. Earnings reported in 1940 are for 1939. In the original, the notes to the table state that the literacy regressions are restricted to ages 16–60 while the appendix (p. 108) gives 15–45. I use the latter. As in the original, all regressions are run without a mean-reversion control (odd columns) and with (even columns); this control is the product of age and a state-level measure of farm-worker wages in 1899 (Lebergott 1964).

The new results provide some reassurance as to the quality of the RC data set replication. As in the original, the coefficients on the treatment terms are generally positive and statistically significant in the earnings regressions and exhibit no consistent pattern in the schooling regressions. However, the match is poorer for adult literacy, for which the replication finds less significance for treatment, at least when including the mean-reversion control.

5.2 Replicating and reanalyzing Figure III and Table VI: Multicensus retrospective cohorts

The last displays in Bleakley (2007), Figure III and Table VI, take the longest view. Unlike Table V, just replicated, they aggregate data from many censuses between 1870 and 1990. Observations are still identified by state and year of birth.

The multicensus RC regressions assess impacts on two IPUMS-provided measures of occupational standing. These are taken to proxy for income, a concept the Census Bureau did not begin directly measuring until the mid-20$^{\text{th}}$ century. Both proxies are constructed from variables that are available for all the rounds used here. The occupational income score (OIS), introduced in section 4.1, is an income index based on reported occupation. Duncan's (1961) socioeconomic index (SEI) blends into the OIS information about educational attainment.

Like Bleakley (2007) Figure II for the SCS regressions, Bleakley (2007) Figure III shows how the cross-sectional association between baseline prevalence and the outcomes of interest varies over time. Bleakley (2007) constructs the figure as follows:

1. The microdata sample is restricted to observations of white males aged 25–60.
2. Within each birth-year cohort between 1825 and 1965, dummies for each census year—or, equivalently, age—are partialled out of the occupational standing indicators.[16]
3. The two occupational standing indicators are then averaged within birth year–birth state–census year cells, producing a three-dimensional panel.
4. Within each birth cohort, the two outcomes are regressed on H while controlling for the 1899 farm wage indicator, a South dummy, and sometimes other state-level variables.

[16] Bleakley (2007) describes the process as partialling out dummies for each census year–birth year combination within the full sample, which is also equivalent.

5. The resulting 141 coefficient estimates for H, $\hat{\beta}_t$, are plotted in Bleakley (2007) Figure III.[17] Then, in Bleakley (2007) Table VI, they are subject to time series analysis to assess whether Exp is a strong explanator. Observations are weighted by the square roots of the cell sizes in step 3.

I make several comments on this methodology, the last of which seems most consequential:

- The census year fixed effects are more properly partialled out of all the regressors, not just the dependent variables. In principle, failure to partial them out of the right-side variables can cause the explanatory power of the fixed effects to load onto the other variables in OLS regressions. In practice, this matters little because the other variables are cross-sectional, and so are nearly orthogonal to the census year effects.[18]
- Aggregating the data before the main analysis prevents controlling for micro-level demographic traits, which the other Bleakley specifications all do. This is relevant if one expands the sample to women and blacks, as I do below. (Age effects are essentially removed by the partialling-out of census year effects within each birth cohort in step 1.)
- While weighting by the square root of cell size is evidently meant to improve efficiency by counteracting heteroskedasticity, theory favors weighting simply by cell size. In general, weights are efficient if inversely proportional to error variance, as in Aitken's generalized least squares estimator. And here the variances of the $\hat{\beta}_t$ can be expected to be inversely proportional to the samples sizes in each cell.
- Three of the five time series specifications reported in Bleakley (2007) Table VI include autoregressive terms: past $\hat{\beta}_t$ are taken as determinants of the current $\hat{\beta}_t$. While this makes for an intuitive robustness test, the specification does not seem grounded in theory. It is hard to see how the cross-sectional association within one birth cohort between historical hookworm burden in the state of birth and future occupational standing would causally affect that association in the next cohort. I make this point less to criticize the autoregressive specifications than to help justify dispensing with them in the reanalysis. The multi-step autoregressive modeling is incompatible with the consolidated OLS approach taken below, which estimates the parameters of interest in a single step.
- The estimation proceeds in three econometric steps—numbers 2, 4, and 5 above—but the imprecision in the first two is not factored into the final one. The time series analysis, though weighting to adjust for the variances of $\hat{\beta}_t$, still conditions on them as if they were observed without error.
- As emphasized in section 3.4, the approach is vulnerable to a spurious regression problem. If the null of no impact from eradication holds, yet long-term convergence occurred, then the regressions could wrongly bestow explanatory power on the eradication campaign. Recognizing the issue, Bleakley (2007) reports time series results that include autoregressive terms as well as polynomial time trends up to order 2. Yet neither tactic obviously suffices. Noise in

[17] Bleakley (2007) symbolizes the coefficients $\hat{\beta}_k$.

[18] They are not exactly orthogonal because the cross-state distribution of the sample varies somewhat from census to census within each birth cohort. The 1920 census, say, could have a higher preponderance of people born in 1890 in historically low-prevalence states than the 1910 census, making H slightly correlated with the census year fixed effects within the 1890 birth cohort.

the data drives the coefficients on the autoregressive terms toward zero. The ambient trend may not be well modeled as quadratic.

Only the penultimate of these concerns was pre-registered. (See section 3.1.)

After reconstructing the original figure and time series regressions, I implement an alternative approach designed to remove or address the above critiques.[19] The alternative starts by copying a practice in the rest of Bleakley (2007), which is to directly fit to microdata. To compute the individual $\hat{\beta}_t$, I fit (2), above, to each birth cohort's microdata. In fact, I consolidate all these regressions into a single, full-sample regression in which the birth year dummies δ_t are interacted with all other right-side variables. This unification facilitates clustering the standard errors by birth state, across cohorts, to adjust for serial correlation.

Then, to explore the ability of the hookworm eradication campaign to explain the $\hat{\beta}_t$, I alter the specification. In particular, I estimate three versions of the full sample model (1), as outlined in section 3.4. The first echoes Bleakley (2007), Table VI, in controlling for polynomial trends in time. With reference to (1), the novel controls inserted in $\mathbf{z}$ are:

$$\{H_j \times t^r\}_{r=0,...,d} \tag{4}$$

where d ranges up to 5 since Bleakley (2007), note 25, reports testing up to that order.

To assess the incremental modeling value of higher-order polynomial terms, I compute and report Schwarz's Bayesian information criterion (BIC) for each fit. For OLS, the BIC is

$$BIC = k \ln N + N(1 + \ln \tau + \ln MSE) \tag{5}$$

where k is the number of modeling parameters; N is sample size; the circle constant τ is twice π; and MSE is the mean squared error. The application of the BIC is complicated here by the assumption that errors are not homoscedastic, and rather are clustered, which implies that the normal likelihood model on which the BIC is based is not accurate. In particular, plugging in the full sizes of the microdata samples for N may misleadingly reward increasing the model parameter count: unsurprisingly, the larger is N the more the BIC will endorse additional parameterization. So, in the spirit of time series modeling in Bleakley (2007), to compute the BIC I view the model as being for the $\hat{\beta}_t$, of which we have 141 observations, from 1825 to 1965. I set N to 141.[20] This choice is conservative from the standpoint of this paper's conclusions, since it raises the bar for adding polynomial terms that may outcompete Exp. For the same reason, I avoid use of the Aikake information criterion, since it puts a smaller penalty on adding terms, replacing $k \ln N$ in (5) with $2k$.

The second model used to study the explanatory power of Exp dispenses with time trend controls. Instead, it introduces three linear spline terms, which generalize Exp to a piecewise-linear contour with two kinks. The kinks occur on the same dates as in Bleakley (2007)'s Exp function,

[19] I initially implemented a bootstrapping approach, in which the combined zeroth and first stages served as the basis for wild bootstrap data generating process. I dropped this after realizing that it could not simulate the AR() processes in the final stage and that for models without AR() terms, the omnibus OLS approach was appropriate, provided it could be made computationally practical.

[20] I set k to the number of parameters in the polynomial model of primary interest, not counting the demographic and other controls. Since this choice is the same for all models, it does not affect the cross-model BIC comparisons.

which are 1910, the nominal campaign start date, and 1891, 19 years before. Since Bleakley (2007) gives Exp a 19-year ramp-up, I give the "before" and "after" segments of the spline—the ones we imagine to be flat—19 years as well. To be precise, the "fixed spline" model replaces $H_j \times Exp_t$ in (1) with three terms:

$$H_j \times t, H_j \times \min(0, t - 1891), \text{and } H_j \times \min(0, t - 1910) \qquad (6)$$

where $\min(\cdot)$ is the minimum function and t is birth year. The sample is restricted to the $3 \times 19 = 57$ birth years between 1872 and 1929. This facilitates testing of whether slope changes occurred in 1891 and 1910 dates, as predicted by a literal reading of the Bleakley (2007) impact model.

A disadvantage of the fixed spline model is that it chooses the kink dates of 1891 and 1910 *a priori*. Yet those dates come from an impact model that, while reasonable, could be inaccurate, as Bleakley (2007) points out. For this reason, the last modeling approach estimates kink points from the data, using a mean squared error criterion, much as in the Bai and Perron (1998) approach to identifying structural breaks.[21] This "flexible spine" model allows exactly two kink dates. The search is exhaustive: all possible pairs of dates are tried when fitting the model to the 1825–1965 data. The method does not easily support formal inference with respect to the kink dates since they are discrete parameters. And, as noted earlier, the model may be drawn to large structural breaks whose timing could not be explained by hookworm eradication. Still, the results are informative as to whether trend shifts in the hookworm eradication era are major features of the historical record.

To start the application, Figure 3 attempts to imitate the original Figure III in data and method. It only departs substantively in adding (95%) confidence intervals for point estimates. Unlike in Bleakley (2007), the Exp step function is not superimposed on the plot. But dashed vertical lines show where it kinks. The original's patterns of dots are recognizable, even if they do not come through exactly.

Following the narrative thread in Bleakley (2007), Table 8, below, seeks to replicate Bleakley (2007) Table VI. It reports time series regressions on the dots in Figure 3. The first row of results is for the SEI regression without full controls, and corresponds to the upper left of Figure 3. The next row is for the bottom-left of Figure 3. And so on. Once more, while the matches are inexact, they are broadly corroborative.

Next, Figure 4 updates Figure 3 by fitting to the expanded data set. Recall from Table 1 that the expansion adds 1860 and 1930 census data and enlarges samples for other years. And, as discussed in section 3.4, it adds blacks and women in order to increase power to detect decade-scale changes. In light of the increased diversity of the sample, all the demographic controls in the Bleakley (2007) single-census RC regressions involving race and sex are included here too.

The data expansion improves the signal-to-noise ratio. The increased stability is obvious from a cursory comparison of Figure 3 and Figure 4, and becomes even clearer after one notes that the vertical ranges on the new graphs are narrower.

[21]The mean-squared error computation factors in sampling weights. The search is constrained to give each segment a length of at least 10 years.

Figure 4 confronts us with the paramount empirical question in the RC analysis: did the association between baseline hookworm prevalence and future occupational standing rise at an historically anomalous rate among the birth cohorts born in the run-up to eradication, between 1891 and 1910? A gaze at Figure 4 suggests that the answer is "no."

To formally test that interpretation, Figure 5, Figure 6, and Figure 7 present the results of fitting the polynomial, fixed spline, and flexible spline models to the expanded microdata. All the figures retain the dots from Figure 4 but drop the grey confidence intervals in order not to obscure the model fits.

First, the polynomial fits examine how controlling flexibly for time trends affects the sign and significance of the coefficient on the treatment proxy $H \times Exp$. In Figure 5, fits of the order-0 through order-5 models are drawn in orange, green, blue, red, purple, and brown, respectively. At the base of the plots, p-values are given for the coefficient on $H \times Exp$, in the same order. In general, models of cubic degree or higher rob the coefficient of much statistical significance, and in some cases reverse its sign. These results are consistent with Bleakley's report that "I have experimented with higher-order polynomial trends and found no estimates of exposure that are statistically significant for $n \leq 5$" (note 25). Yet the BIC favors cubic or quartic fits, as shown in the corresponding Table 9, where BIC-preferred results are bolded.

The "fixed spline" model fits also do not support the hypothesis that the hookworm eradication campaign affected future occupational standing. Figure 6 shows these results. As in the earlier SCS graphs, red lines plot the trends of particular interest while p-values for slope changes at the two chosen kink points are reported beneath. Across the four panes of the figure, the p-values for a slope change at the first kink, between 1872–91 and 1891–1910, range between 0.33 and 0.69. For the second kink, they range between 0.18 and 0.86, with the lowest p-value associated with another bend in the theoretically wrong direction.

Finally, Figure 7 lets the data choose the two most important kink points in each 141-year time series. None produces a pattern convincingly similar to Exp. In a robustness test, allowing four kinks instead of two does not reverse this impression. (See Figure A3.) This does not prove that the hookworm campaign left no imprint on the series in question. It does suggest that any imprint is too modest to leave a compelling fingerprint. And it highlights the possibility that other factors are at work over the long term that could be spuriously attributed to hookworm eradication.

The divergence in conclusion from Bleakley (2007) is not an artifact of adding blacks and women to the multicensus RC samples. In the appendix, Table A1, Figure A4, Figure A5, and Figure A6 repeat the foregoing exercise while restricting to white men. They too fail to generate evidence for long-term impacts of the hookworm eradication campaign on occupational standing.

6 *Conclusion*

Bleakley (2007) identifies impacts from a variable that results from the interaction of two factors: the geographic pattern of baseline hookworm burden and the timing the eradication campaign. The first factor is not credibly exogenous since it is a marker for climate and geography, thus economic history. The second can be taken as more exogenous, at least in the short term. Thus, given the priors I bring to this study, for it to produce strong evidence of impact from the campaign, it must demonstrate historically anomalous changes in the outcomes of interest in the time dimension, and that over a range measured in years rather than decades. And it must do so while credibly warding off the possibility of spurious attribution of unrelated long-term dynamics.

In my view, none of the regressions in Bleakley (2007) specifies the model richly enough in the time domain to produce such evidence. Most of them effectively fit to a step function while controlling linearly for time. These models can easily generate misleading results when fit to a series with long-term structure such as an S curve. The graphs in Bleakley (2007) appear to belie this concern by demonstrating to the naked eye that the time series of interest are well modeled by step functions. But those results appear fragile, especially to sample expansion.

Most of the revisions and tests on which I base the judgment of fragility were not pre-registered. One exceptions is the set of tests controlling for polynomials up to order 5, which were implicitly pre-registered since Bleakley too ran such regressions. That said, most of the innovations spring from relatively natural sources: using the latest data sets from IPUMS, and copying choices from specification to specification within the original paper.

Without access to the original data and code, we cannot determine to what extent the discrepancies in the replication owe to errors in either version, to subtle differences in variable construction, or to IPUMS revisions. However, the full data and code for this replication are posted. Unless the original data and code become accessible, I believe that this new version should be taken as the reference implementation of Bleakley (2007). Only it can be subject to the review and replication that characterize science.

7 References

Bai, Jushan, and Pierre Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66 (1): 47–78. DOI: 10.2307/2998540.

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): 1637–80. DOI: 10.1093/qje/qjw022.

Bleakley, Hoyt (2002). "Disease and Development: Evidence from Hookworm Eradication in the American South." Job market paper. Available at j-bradford-delong.net/movable_type/refs/Mozilla_Scrapbook3/Hookworm_October8.pdf.

Bleakley, Hoyt (2006). "Disease and Development: Evidence from Hookworm Eradication in the American South." Working Paper 205. Stigler Center, University of Chicago.

Bleakley, Hoyt (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South." *The Quarterly Journal of Economics* 122 (1): 73–117. DOI: 10.1162/qjec.121.1.73.

Bleakley, Hoyt (2009). "Disease and Development: Evidence from Hookworm Eradication in the American South." In Michael Spence and Maureen Lewis, eds. *Health and Growth: Commission on Growth and Development*. World Bank. DOI: 10.1596/978-0-8213-7659-1.

Bleakley, Hoyt (2010). "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics* 2 (2): 1–45. DOI: 10.1257/app.2.2.1.t.

Bleakley, Hoyt (2018). "Corrigenda for 'Disease and Development: Evidence from Hookworm Eradication in the American South". January 5. Available at umich.edu/~hoytb/Bleakley_Hookworm_Corrigenda.pdf.

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

Bureau of the Census (1915). *Wealth, Debt and Taxation. Vol. II*. Washington, DC: U.S. Government Printing Office.

Bureau of the Census (1935). *Financial Statistics of State and Local Governments: 1932*. Washington, DC: U.S. Government Printing Office.

Croke, Kevin (2014). The Long Run Effects of Early Childhood Deworming on Literacy and Numeracy: Evidence from Uganda. *Unpublished Manuscript*. Available at scholar.harvard.edu/files/kcroke/files/ug_lr_deworming_071714.pdf.

Croke, Kevin, Joan Hamory Hicks, Eric Hsu, Michael Kremer, and Edward Miguel (2016). "Does Mass Deworming Affect Child Nutrition? Meta-Analysis, Cost-Effectiveness, and Statistical Power." *Working Paper Series*. National Bureau of Economic Research. DOI: 10.3386/w22382.

Duncan, Otis Dudley (1961). *A Socioeconomic Index for All Occupations*. In Occupations and Social Status, edited by Reiss, Albert J. , 109–38. New York: Free Press.

Inter-university Consortium for Political and Social Research (ICPSR). Historical, Demographic, Economic, and Social Data: The United States, 1790–1970. Ann Arbor, MI. DOI: 10.3886/ICPSR02896.v3.

Kofoid, Charles A., and John P. Tucker (1921). "On the Relationship of Infection by Hookworm to the Incidence of Morbidity and Mortality in 22,842 Men of the United States Army," *American Journal of Hygiene*, 1, 79-117. Available at babel.hathitrust.org/cgi/pt?id=uc1.\protect\T1\textdollarc46889;view=1up;seq=3.

Lebergott, Stanley. *Manpower in Economic Growth: The American Record Since 1800*. New York, NY: McGraw-Hill, 1964.

Linnemer, Laurent, and Michael Visser (2017). "The Most Cited Articles from the Top-5 Journals (1991–2015)." Available at hal.archives-ouvertes.fr/hal-01634432.

Miguel, Edward, and Michael Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217. DOI: 10.1111/j.1468-0262.2004.00481.x.

Ozier, Owen (2018). "Exploiting Externalities to Estimate the Long-Term Effects of Childhood Deworming." *American Economic Journal: Applied Economics* 10 (3): 235–62. DOI: 10.1257/app.20160183.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek (2015) *Integrated Public Use Microdata Series: Version 6.0 [dataset]*. Minneapolis: University of Minnesota. DOI: 10.18128/D010.V6.0.

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*, No. 7, CD000371. DOI: 10.1002/14651858.CD000371.pub6.

Welch, Vivian A., Elizabeth Ghogomu, Alomgir Hossain, Shally Awasthi, Zulfiqar A. Bhutta, Chisa Cumberbatch, Robert Fletcher, et al. (2017). "Mass Deworming to Improve Developmental Health and Wellbeing of Children in Low-Income and Middle-Income Countries: A Systematic Review and Network Meta-Analysis." *The Lancet Global Health* 5 (1): e40–50. DOI: 10.1016/S2214-109X(16)30242-X.

Table 1: IPUMS Census Samples in Original and Expanded Data Sets

Census year	Original (estimated)	Expanded
1860	0%	1.2%[1]
1870	1%	1.2%
1880	1%/100%[2]	10%/100%[2]
1890	0%	0%
1900	0.13%/1%[2]	5%
1910	0.4%	100%[3]
1920	1%	100%[3]
1930	0%	100%
1940	1%	100%
1950[4]	0.2%/1%[2]	0.2%/1%[2]
1960	1%	5%
1970	1%	1%
1980	5%	5%
1990	5%	5%

[1]Excludes slaves.
[2]Pairs of numbers refer separately to SCS and RC regressions.
[3]For SCS literacy regressions, 1910 1.4% and 1920 1% samples used instead.
[4]Data source is same for SCS and RC, but SCS dependent variables (on schooling) only observed in a fifth of cases.

Table 2: Displays in Bleakley (2007)

Display	Research design	Unit of observation	Demographic groups	Outcomes	Tested with full controls?
Tables II & III	SCS	Individual	Blacks & whites, men & women	In school, in school full-time, literate	Yes
Figure II	SCS	Individual	Blacks & whites, men & women	In school	No
Table IV	SCS	Individual	Blacks & whites, men & women	Literate, in labor force, occupational standing, lives in city	No
Table V	RC	Individual	Blacks & whites, men & women	Earnings, years of schooling, literate	No
Figure III, Table VI	RC	Birth year–birth state	White men	Occupational income score, Duncan's socioeconomic index	Yes

Table 3: Replication of Bleakley (2007) Table I: Summary statistics

	Whole Sample		By Hookworm Infection >40%		By Hookworm Infection <40%	
	Original	New	Original	New	Original	New
Hookworm-Infection Rate	0.320 (0.230)	0.333 (0.226)	0.554 (0.137)	0.559 (0.135)	0.164 (0.117)	0.176 (0.117)
Individuals Treated At Least Once by the RSC, Per School-Age Child	0.206 (0.205)	0.103 (0.181)	0.342 (0.199)	0.190 (0.247)	0.109 (0.147)	0.043 (0.068)
School Enrollment, 1910	0.721 (0.104)	0.721 (0.103)	0.711 (0.099)	0.708 (0.100)	0.729 (0.108)	0.729 (0.105)
Change in School Enrollment, 1910–1920	0.089 (0.080)	0.130 (0.077)	0.103 (0.090)	0.147 (0.088)	0.078 (0.072)	0.118 (0.067)
Full-time School Attendance, 1910	0.517 (0.140)	0.519 (0.139)	0.469 (0.123)	0.481 (0.135)	0.551 (0.141)	0.546 (0.136)
Change in Full-time School Attendance, 1910–1920	0.203 (0.097)	0.235 (0.095)	0.246 (0.093)	0.268 (0.103)	0.172 (0.089)	0.212 (0.083)
Literacy, 1910	0.853 (0.104)	0.853 (0.105)	0.824 (0.101)	0.822 (0.103)	0.875 (0.102)	0.874 (0.102)
Change in Literacy, 1910–1920	0.060 (0.067)	0.061 (0.069)	0.081 (0.075)	0.084 (0.079)	0.045 (0.057)	0.045 (0.056)
Population Black, 1910	0.357 (0.221)	0.342 (0.200)	0.410 (0.208)	0.406 (0.171)	0.318 (0.223)	0.298 (0.207)
Fraction Population Urban, 1910	0.174 (0.200)	0.177 (0.191)	0.167 (0.214)	0.172 (0.197)	0.180 (0.223)	0.180 (0.188)
School term, in Months, c. 1910	5.251 (1.066)	5.490 (0.902)	5.055 (1.042)	5.191 (0.705)	5.391 (1.068)	5.698 (0.968)
School per Square Mile, c. 1910	0.195 (0.358)	0.143 (0.055)	0.142 (0.053)	0.125 (0.039)	0.233 (0.465)	0.155 (0.062)
Value of School Property, per Pupil, Current Dollars, c. 1910	5.518 (4.037)	6.913 (6.526)	4.699 (3.159)	5.402 (3.925)	6.104 (4.496)	7.943 (7.678)
Teacher-to-School Ratio, c. 1910	1.336 (0.545)	1.394 (0.479)	1.397 (0.505)	1.334 (0.390)	1.293 (0.572)	1.436 (0.530)
Sample size	115	117	48	48	67	69

Variable means displayed with standard deviations in parentheses beneath. "Original" results copied from Bleakley (2007) Table I. "New" results computed after reconstructing the data set from primary sources listed in Bleakley (2007) appendices.

Table 4: Replication of Bleakley (2007) Table II: Hookworm and human capital: Basic results

		School enrollment		Full-time school attendance		Literacy	
		Original	New	Original	New	Original	New
Panel A: Basic results							
Census years	Include SEA-specific time trends?						
1910–1920	No	0.0883***	0.0986***	0.1591***	0.1670***	0.0587***	0.0675***
		(0.0225)	(0.0223)	(0.0252)	(0.0243)	(0.0186)	(0.0174)
Observations		64676	65396		65396	49476	50028
1900–1950	No	0.0608**	0.0724***	0.1247***	0.1188***		
		(0.0261)	(0.0230)	(0.0286)	(0.0237)		
Observations		140161	141329		141329		
1900–1950	Yes	0.0954***	0.1087***	0.1471***	0.1618***		
		(0.0233)	(0.0294)	(0.0287)	(0.0346)		
Observations		140161	141329		141329		
Panel B: Effects within and between states							
Include state x Post dummies		0.1313***	0.1231***	0.2144***	0.2050***	0.0417**	0.0511***
		(0.0245)	(0.0338)	(0.0290)	(0.0342)	(0.0207)	(0.0188)
Observations			141329		141329		50028
Allow for state-specific mean reversion		0.1148***	0.1103***	0.1813***	0.1962***	0.0408**	0.0200
		(0.0265)	(0.0407)	(0.0312)	(0.0364)	(0.0206)	(0.0190)
Observations			141329		141329		50028
Use infection from state of birth instead of SEA		0.0489	0.0712	0.2057***	0.0931	0.0907**	–0.0437
		(0.0504)	(0.0738)	(0.0765)	(0.1556)	(0.0451)	(0.0417)
Observations			665263		665263		185943
Census years		1900–1950		1900–1950		1910–1920	
Include SEA-specific time trends?		Yes		Yes		No	

"Original" results copied from Bleakley (2007) Table II. "New" results computed after reconstructing the data set from primary sources. New regressions weighted by IPUMS-provided sampling weights. Standard errors in parentheses, clustered by state economic area, except in the last row, where they are clustered by state. Where reported, sample sizes from original are from Bleakley (2002), Tables III & VI. Standard errors in parentheses, clustered by state economic area, except in the last row, where they are clustered by state. $^{*}p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

Table 5: Replication of Bleakley (2007) Table III: Sensitivity tests and results for subgroups

	School enrollment, 1900–50		School enrollment, 1910–20		Full-time school attendance, 1900–50		Full-time school attendance, 1910–20		Literacy, 1910–20	
	Original	New	Original	New	Original	New	Original	New	Original	New
Panel A: Baseline results										
Baseline	0.0954***	0.1087***	0.0883***	0.0986***	0.1471***	0.1618***	0.1591***	0.1670***	0.0587***	0.0675***
	(0.0233)	(0.0294)	(0.0225)	(0.0223)	(0.0287)	(0.0346)	(0.0252)	(0.0243)	(0.0186)	(0.0174)
Observations	140161	141329	64676	65396		141329		65396	49476	50028
Panel B: Specifications with additional controls										
Health & health policy	0.1200***	0.1146***	0.1187***	0.1162***	0.1628***	0.1540***	0.1646***	0.1423***	0.0724***	0.0679***
	(0.0291)	(0.0371)	(0.0262)	(0.0223)	(0.0355)	(0.0374)	(0.0294)	(0.0242)	(0.0233)	(0.0212)
Observations		131662		61285		131662		61285		46894
Education & race	0.1235***	0.0999***	0.0793***	0.0548***	0.1851***	0.1684***	0.1581***	0.1317***	0.0556***	0.0377**
	(0.0208)	(0.0275)	(0.0208)	(0.0169)	(0.0247)	(0.0308)	(0.0250)	(0.0217)	(0.0171)	(0.0162)
Observations		140154		64900		140154		64900		49645
Full controls	0.1014***	0.0256	0.0850***	0.0029	0.1408***	0.0909**	0.1026***	0.0553**	0.0513**	–0.0160
	(0.0349)	(0.0416)	(0.0224)	(0.0221)	(0.0421)	(0.0393)	(0.0325)	(0.0229)	(0.0213)	(0.0220)
Observations		131062		61027		131062		61027		46695
Panel C: Demographic subgroups										
Preteens	0.0932***	0.0972***	0.0890***	0.1015***	0.1416***	0.1458***	0.1549***	0.1679***	0.0912***	0.0978***
	(0.0255)	(0.0318)	(0.0242)	(0.0227)	(0.0302)	(0.0361)	(0.0266)	(0.0242)	(0.0253)	(0.0231)
Observations		80711		38007		80711		38007		22639
Adolescents	0.0986***	0.1312***	0.0877***	0.0977***	0.1573***	0.1905***	0.1682***	0.1694***	0.0323*	0.0438***
	(0.0280)	(0.0369)	(0.0282)	(0.0284)	(0.0336)	(0.0434)	(0.0295)	(0.0295)	(0.0165)	(0.0167)
Observations		60618		27389		60618		27389		27389
Blacks	0.2299***	0.1819***	0.1838***	0.1612***	0.2601***	0.2207***	0.2205***	0.1956***	0.1078***	0.1197***
	(0.0399)	(0.0533)	(0.0337)	(0.0335)	(0.0399)	(0.0550)	(0.0320)	(0.0328)	(0.0374)	(0.0361)
Observations		46464		22824		46464		22824		17528
Whites	0.0378	0.0878***	0.0270	0.0553**	0.1103***	0.1589***	0.1169***	0.1419***	0.0264*	0.0284**
	(0.0237)	(0.0306)	(0.0267)	(0.0269)	(0.0294)	(0.0339)	(0.0294)	(0.0275)	(0.0139)	(0.0129)
Observations		94865		42572		94865		42572		32500

"Original" results copied from Bleakley (2007) Table III. "New" results computed after reconstructing the data set from primary sources listed in Bleakley (2007) appendices. 1900–50 regressions include SEA-specific time trends, in accordance with the original's equation 2. 1910–20 regressions do not, in accordance with the original's equation 1. New regressions weighted by IPUMS-provided sampling weights. Standard errors in parentheses, clustered by state economic area. Where reported, sample sizes from original are from Bleakley (2002), Tables III & VI. Standard errors in parentheses, clustered by state economic area. $*p < 0.1$. $**p < 0.05$. $***p < 0.01$.

Table 6: Replication of Bleakley (2007) Table IV: Contemporaneous effect on adult outcomes

Sample:	Whole		Male		Female		White		Black	
	Original	New	Original	New	Original	New	Original	New	Original	New
Literacy	0.0062	0.0266***	–0.0107	0.0171*	0.0203	0.0374***	0.0107	0.0132	–0.0014	0.0450***
	(0.0095)	(0.0077)	(0.0108)	(0.0098)	(0.0127)	(0.0097)	(0.0112)	(0.0100)	(0.0229)	(0.0165)
Observations	97187	98373		49530		48843		65763		32610
Labor-force participation	–0.0069	–0.0106	–0.0069	–0.0075	–0.0056	–0.0095	–0.0212*	–0.0271**	0.0036	0.0160
	(0.0134)	(0.0111)	(0.0065)	(0.0057)	(0.0284)	(0.0256)	(0.0124)	(0.0104)	(0.0249)	(0.0196)
Observations	97778	98373		49530		48843		65763		32610
Occupational income score	0.0526	0.2633	–0.0186	0.4569	0.0581	–0.3106	0.0855	0.6591	0.0224	–0.2146
	(0.2836)	(0.3234)	(0.4912)	(0.3870)	(0.4163)	(0.5248)	(0.3903)	(0.5319)	(0.3861)	(0.2937)
Observations		60816		48289		12527		37024		23792
Lives in urban area	0.0157	0.0072	0.0030	0.0031	0.0280	0.0111	0.0199	0.0019	0.0132	0.0182
	(0.0172)	(0.0119)	(0.0190)	(0.0143)	(0.0177)	(0.0131)	(0.0226)	(0.0172)	(0.0245)	(0.0194)
Observations		98373		49530		48843		65763		32610

"Original" results copied from Bleakley (2007) Table IV. "New" results computed after reconstructing the data set from primary sources listed in Bleakley (2007) appendices. New regressions weighted by IPUMS-provided sampling weights. Where reported, sample sizes from original are from Bleakley (2002), Table VII. Standard errors in parentheses, clustered by state economic area. *$p < 0.1$. **$p < 0.05$. ***$p < 0.01$.

Table 7: Replication of Bleakley (2007) Table V: Long-term follow-up based on intensity of exposure to the treatment campaign

Dependent variables:	Log earnings, 1939				Years of schooling, 1940				Literacy status, 1920			
Control for mean reversion:	No		Yes		No		Yes		No		Yes	
	Original	New	Original	New	Original	New	Original	New	Original	New	Original	New
Panel A: Main results												
Independent variables												
Hookworm infection rate × Years of exposure	0.0286***	0.0154***	0.0234*	0.0197**	–0.0243	–0.0119	0.0037	0.0326	0.0158***	0.0065***	0.0115***	0.0028
	(0.0066)	(0.0056)	(0.0093)	(0.0092)	(0.0328)	(0.0276)	(0.0357)	(0.0380)	(0.0019)	(0.0017)	(0.0020)	(0.0024)
Observations		257525		256806		537272		536029		407171		406200
Panel B: Changing returns to schooling												
Independent variables												
Hookworm infection rate × Years of exposure	0.0254***	0.0161***	0.0219***	0.0189***								
	(0.0044)	(0.0029)	(0.0063)	(0.0049)								
Hookworm infection rate × Years of exposure × Years of schooling	0.0023**	0.0024***	0.0022**	0.0024***								
	(0.0009)	(0.0007)	(0.0009)	(0.0008)								
Observations		257525		256806								
Panel C: Estimates of hookworm × exposure for demographic subgroups												
Subsamples												
Males	0.0265***	0.0119**	0.0253***	0.0207**	–0.0690**	–0.0492*	–0.0376	–0.0035	0.0108***	0.0010	0.0083***	–0.0034*
	(0.0056)	(0.0049)	(0.0080)	(0.0086)	(0.0326)	(0.0263)	(0.0347)	(0.0361)	(0.0018)	(0.0015)	(0.0019)	(0.0019)
Observations		189936		189491		266844		266275		201776		201344
Females	0.0322***	0.0259**	0.0157	0.0168	0.0200	0.0250	0.0444	0.0684	0.0209***	0.0118***	0.0148***	0.0087**
	(0.0115)	(0.0111)	(0.0165)	(0.0159)	(0.0338)	(0.0296)	(0.0385)	(0.0435)	(0.0027)	(0.0022)	(0.0030)	(0.0033)
Observations		67589		67315		270428		269754		205395		204856
Whites	0.0293***	0.0153***	0.0232**	0.0186*	–0.0110	–0.0008	0.0164	0.0436	0.0131***	0.0048***	0.0086***	0.0002
	(0.0071)	(0.0057)	(0.0103)	(0.0103)	(0.0345)	(0.0282)	(0.0378)	(0.0392)	(0.0022)	(0.0014)	(0.0020)	(0.0018)
Observations		227863		227359		480376		479501		358048		357414
Blacks	0.0220***	0.0159*	0.0253**	0.0289***	0.1013***	–0.0799**	0.0133	0.0253	0.0314***	0.0147***	0.0262***	0.0119*
	(0.0072)	(0.0086)	(0.0103)	(0.0099)	(0.0387)	(0.0371)	(0.0461)	(0.0561)	(0.0065)	(0.0048)	(0.0063)	(0.0064)
Observations		29662		29447		56896		56528		49123		48786

"Original" results copied from Bleakley (2007) Table V. "New" results computed after reconstructing the data set from primary sources. New regressions weighted by IPUMS-provided sampling weights. In panels A and C, each cell holds results from a different regression, whereas in panel B, each column does. Earnings and schooling regressions restricted to ages 25–60. Literacy regressions restricted to ages 15–45. Standard errors in parentheses, clustered by state of birth. $^{*}p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

Table 8: Replication of Bleakley (2007) Table VI: Exposure to RSC versus alternative time-series relationships

Bleakley (2007) specification:		(1)		(2)		(3)		(4)		(5)	
Outcome	Controls	Original	New	Original	New	Original	New	Original	New	Original	New
Duncan's socioeconomic indicator	Basic	0.5352*** (0.0418)	0.3669*** (0.0417)	0.7566*** (0.1069)	0.4599*** (0.0882)	0.3928*** (0.0520)	0.3343*** (0.0566)	0.5983*** (0.1124)	0.4239*** (0.0911)	0.4858*** (0.1282)	0.3447*** (0.0992)
Duncan's socioeconomic indicator	Full controls	0.5007*** (0.0661)	0.7676*** (0.1038)	0.8820*** (0.1707)	0.9784*** (0.2344)	0.3544*** (0.0735)	0.9162*** (0.1338)	0.6616*** (0.1791)	1.1660*** (0.2686)	0.7081*** (0.1969)	1.2994*** (0.2871)
Occupational income score	Basic	0.3113*** (0.0214)	0.2836*** (0.0230)	0.2915*** (0.0542)	0.1843*** (0.0549)	0.2612*** (0.0384)	0.2788*** (0.0398)	0.2497*** (0.0612)	0.1786*** (0.0550)	0.1912*** (0.0622)	0.1322*** (0.0482)
Occupational income score	Full controls	0.2623*** (0.0339)	0.3485*** (0.0525)	0.3732*** (0.0858)	0.3164*** (0.1135)	0.2346*** (0.0438)	0.3959*** (0.0547)	0.3393*** (0.0960)	0.3551*** (0.1215)	0.2742*** (0.1007)	0.3375*** (0.1181)
Order of Polynomial Trend:		0		1		0		1		2	
Order of Autoregressive Process:		0		0		1		1		2	

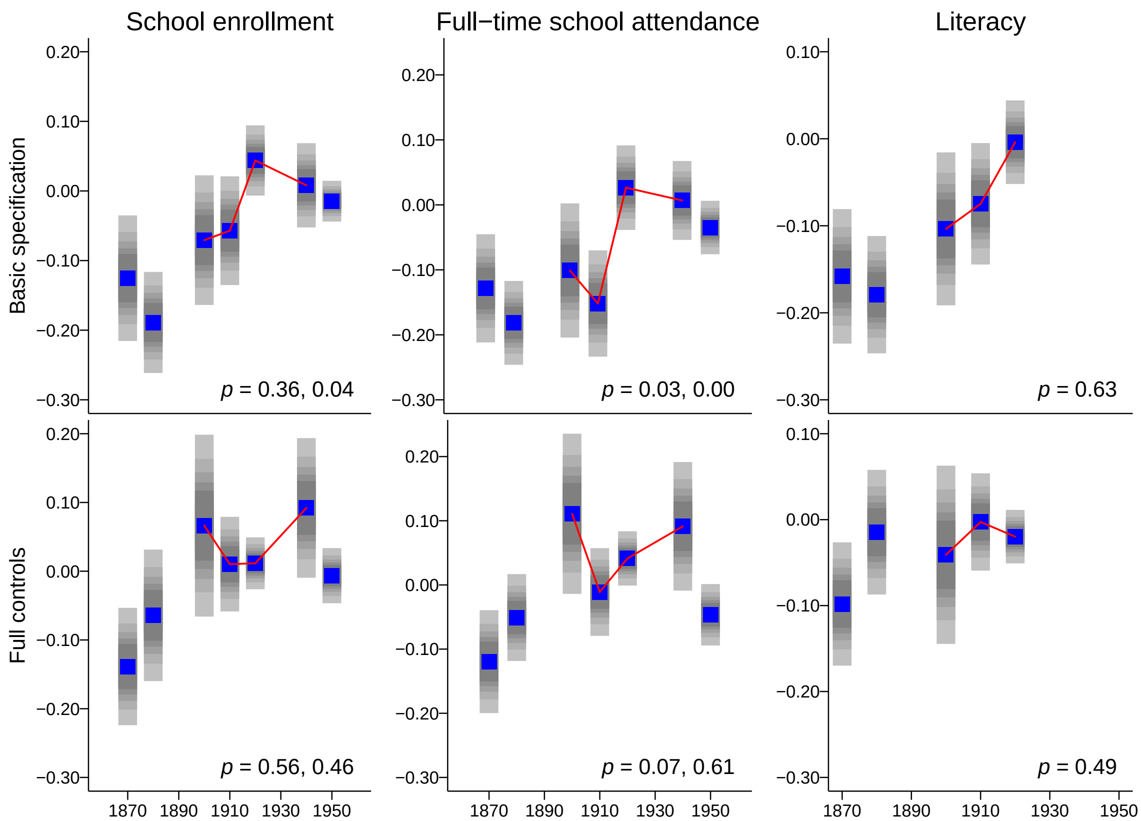
"Original" results copied from Bleakley (2007) Table VI. "New" results computed after reconstructing the data set from primary sources. Rows are in a different order than in the original. New regressions weighted by IPUMS-provided sampling weights. Heteroskedasticity-robust standard errors in parentheses. $^{***}p < 0.01$.

Table 9: Revision of Bleakley (2007) Table VI: Exposure to RSC versus alternative time-series relationships

Outcome	Controls	Coefficient on $H \times Exp$					
Duncan's socioeconomic indicator BIC	Basic	0.0940 (0.1426) 1253.83	0.1020 (0.1707) 1259.37	0.0717 (0.1522) 1259.50	–0.2129 (0.1547) 1222.53	**–0.1918** **(0.1369)** **1219.79**	–0.3579** (0.1386) 1226.41
Duncan's socioeconomic indicator BIC	Full controls	0.3595 (0.2655) 1506.05	0.6697*** (0.2073) 1518.33	0.4794** (0.1875) 1498.48	**0.1667** **(0.1856)** **1493.21**	0.1516 (0.1828) 1499.42	–0.3911 (0.2460) 1508.30
Occupational Income Score BIC	Basic	0.1997* (0.1128) 1058.04	0.1179 (0.0937) 1042.71	0.1329 (0.1077) 1034.89	0.0039 (0.0680) 998.35	**0.0153** **(0.0747)** **994.19**	–0.0441 (0.0632) 997.93
Occupational Income Score BIC	Full controls	0.1890*** (0.0684) 1275.41	0.2106** (0.0791) 1279.82	0.1968* (0.1004) 1286.56	**0.0176** **(0.0917)** **1268.69**	–0.0045 (0.0931) 1275.62	–0.1828** (0.0894) 1269.80
Order of Polynomial Trend:		0	1	2	3	4	5

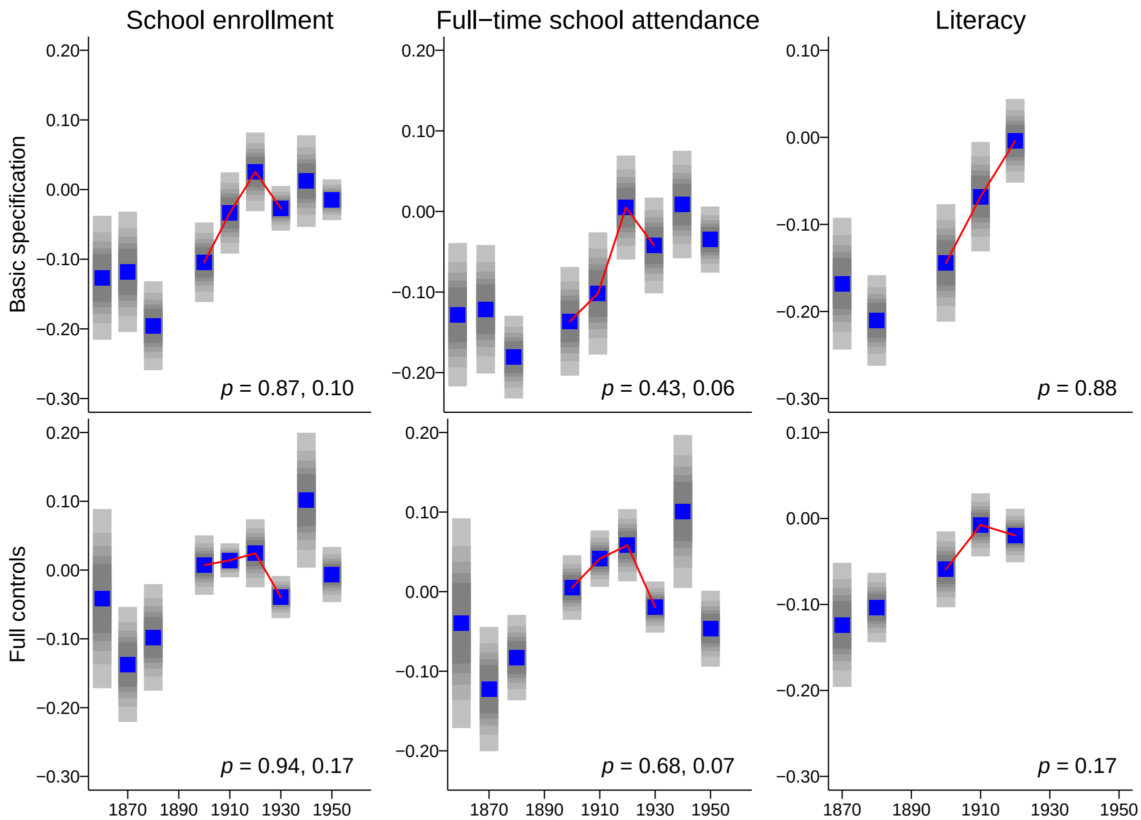
Estimates based on expanded data set, including blacks as well as whites. Regressions weighted by IPUMS-provided sampling weights. Standard errors clustered by state of birth in parentheses. BIC is the Bayesian Information Criterion, taking sample size as 141 and mean-squared error from data points and model fits presented in Figure 5, weighting by number of observations in the cell for each data point. Bolded results in each row are those favored by the BIC. $*p < 0.1$. $**p < 0.05$. $***p < 0.01$.

Figure 1: Replication and extension of Bleakley (2007) Figure II



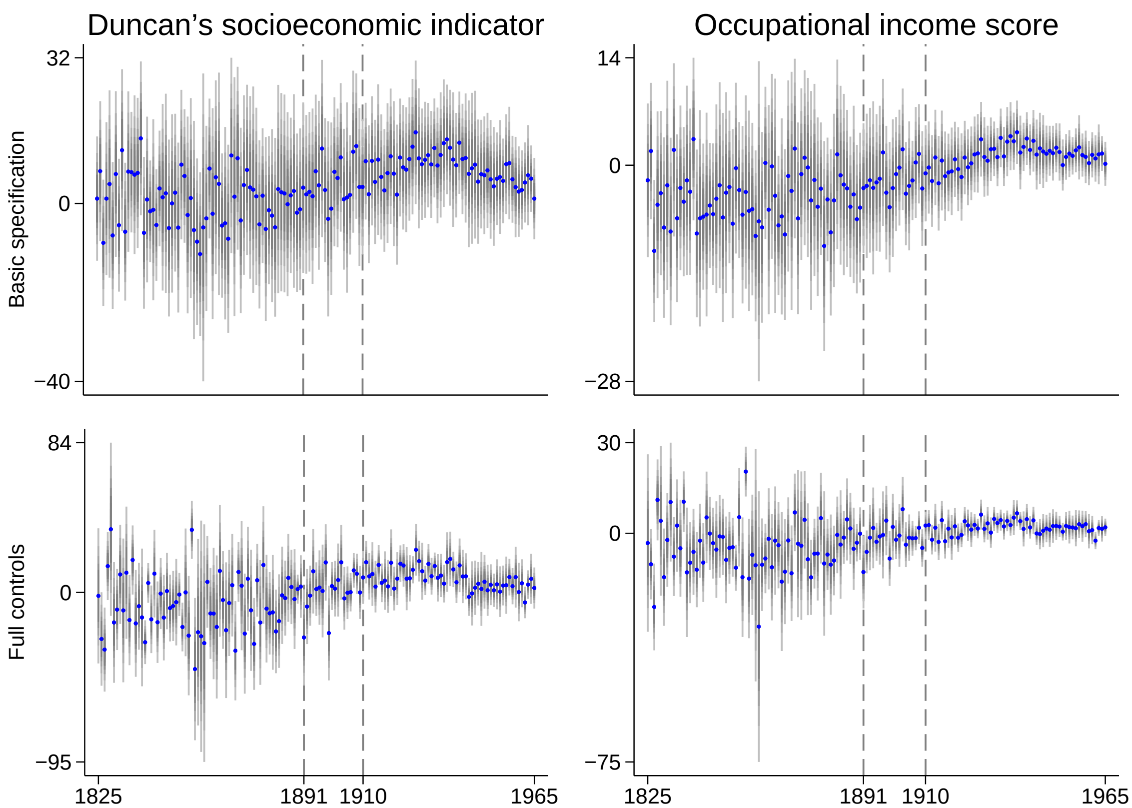
Notes: Blue dots depict point estimates of the cross-SEA association in each census round between baseline hookworm prevalence and the outcome listed. Grey bars show 95% confidence intervals. Red contours highlight the quantities of particular interest, the rates of change just before, during, and just after 1910–20. p-values in each pane are from two-tailed tests for kinks in the red contours, the first at 1910, the second at 1920. Lack of literacy data after 1920 prevents the second test.

Figure 2: Replication and extension of Bleakley (2007) Figure II: Expanded data set



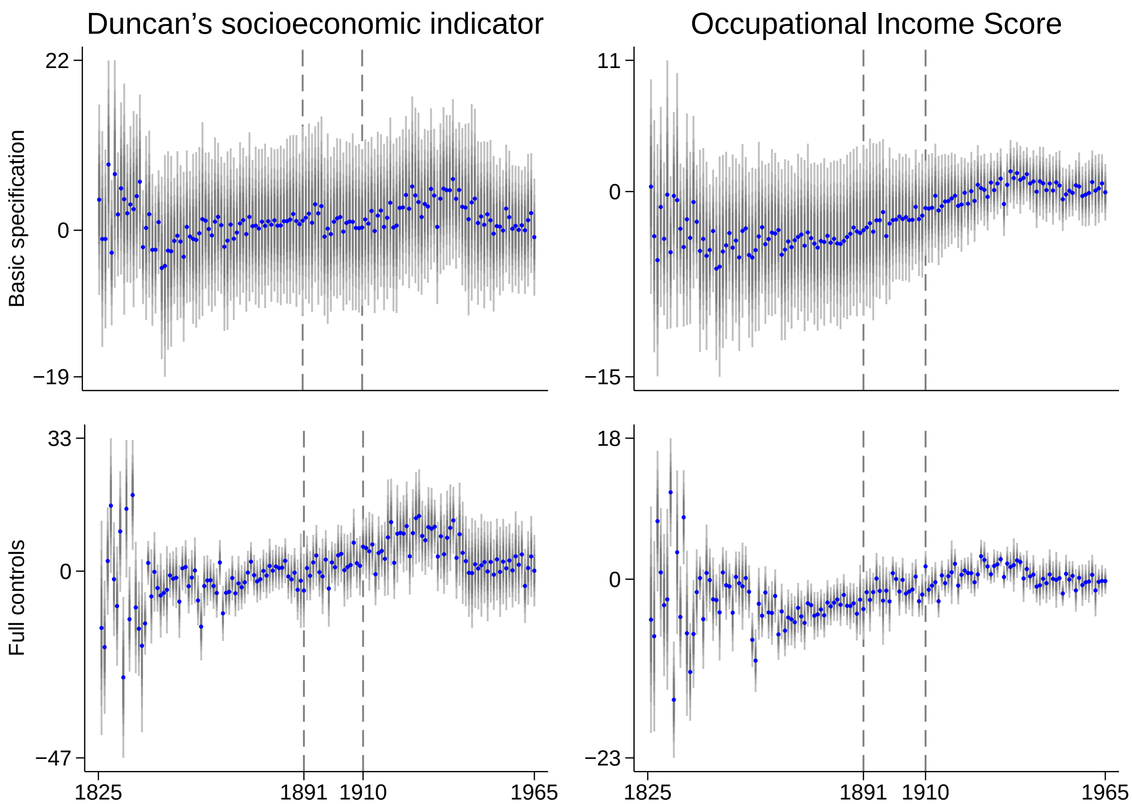
Notes: Blue dots depict point estimates of the cross-SEA association in each census round between baseline hookworm prevalence and the outcome listed. Grey bars show 95% confidence intervals. Red contours highlight the quantities of particular interest, the rates of change just before, during, and just after 1910–20. p-values in each pane are from two-tailed tests for kinks in the red contours, the first at 1910, the second at 1920. Lack of literacy data after 1920 prevents the second test.

Figure 3: Replication and extension of Bleakley (2007) Figure III: Reconstructed data set



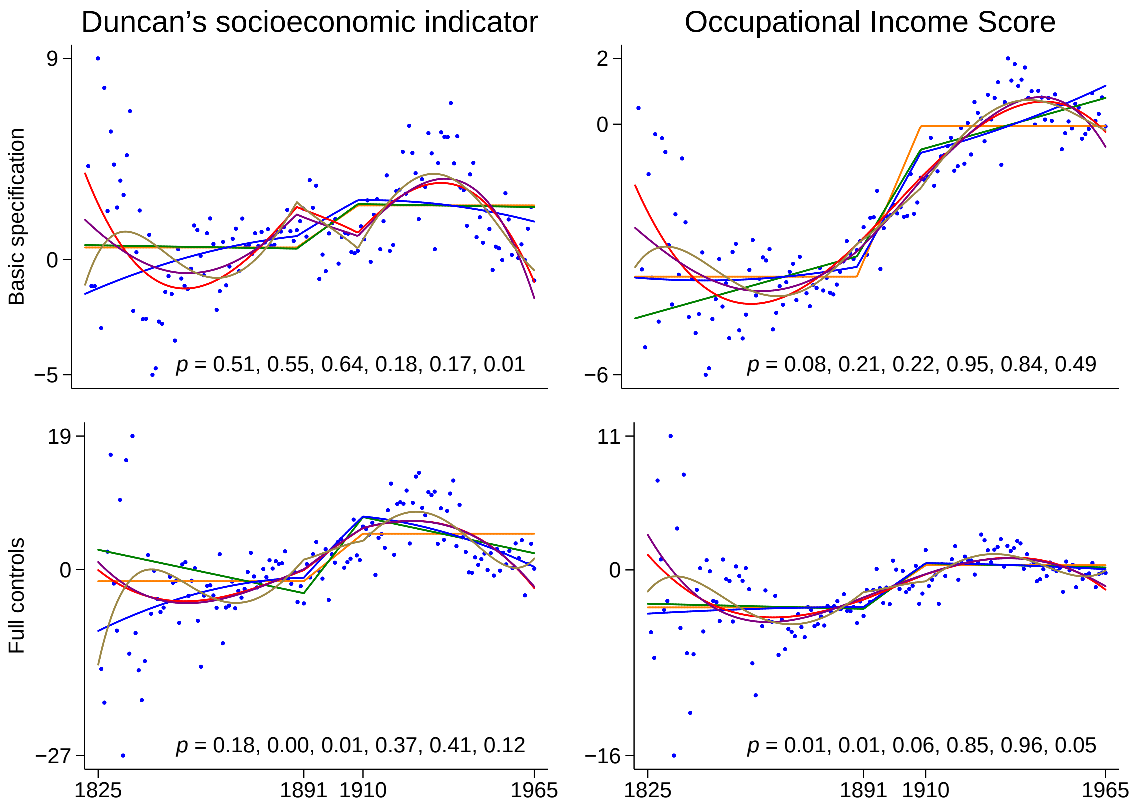
Notes: Blue dots depict point estimates of the cross-state association of baseline hookworm prevalence with the outcome shown within each birth cohort. Grey bars show 95% confidence intervals. Vertical grey lines indicate kink points in Bleakley (2007) exposure function, Exp, which bends upward at the first and plateaus at the second.

Figure 4: Replication and extension of Bleakley (2007) Figure III: Expanded data set



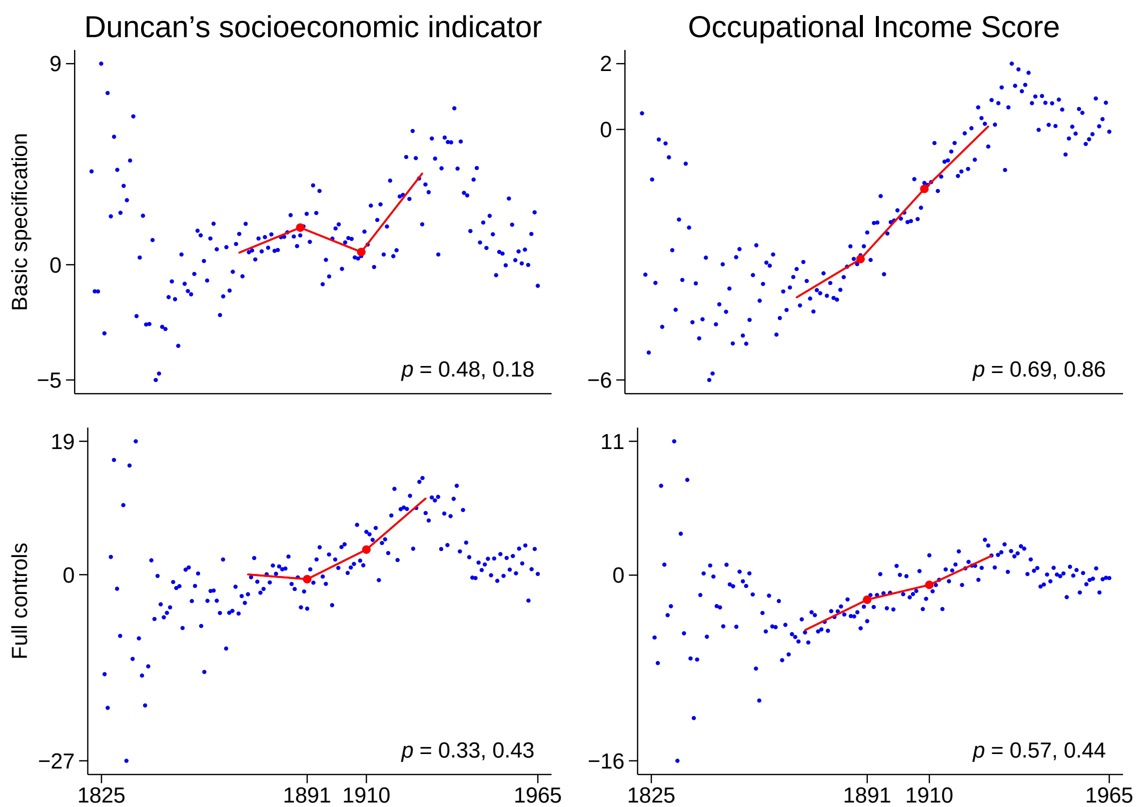
Notes: Blue dots depict point estimates of the cross-state association of baseline hookworm prevalence with the outcome shown within each birth cohort. Grey bars show 95% confidence intervals. Vertical grey lines indicate kink points in Bleakley (2007) exposure function, Exp, which bends upward at the first and plateaus at the second.

Figure 5: Replication and extension of Bleakley (2007) Figure III: Model with polynomial time controls, fit to expanded data set



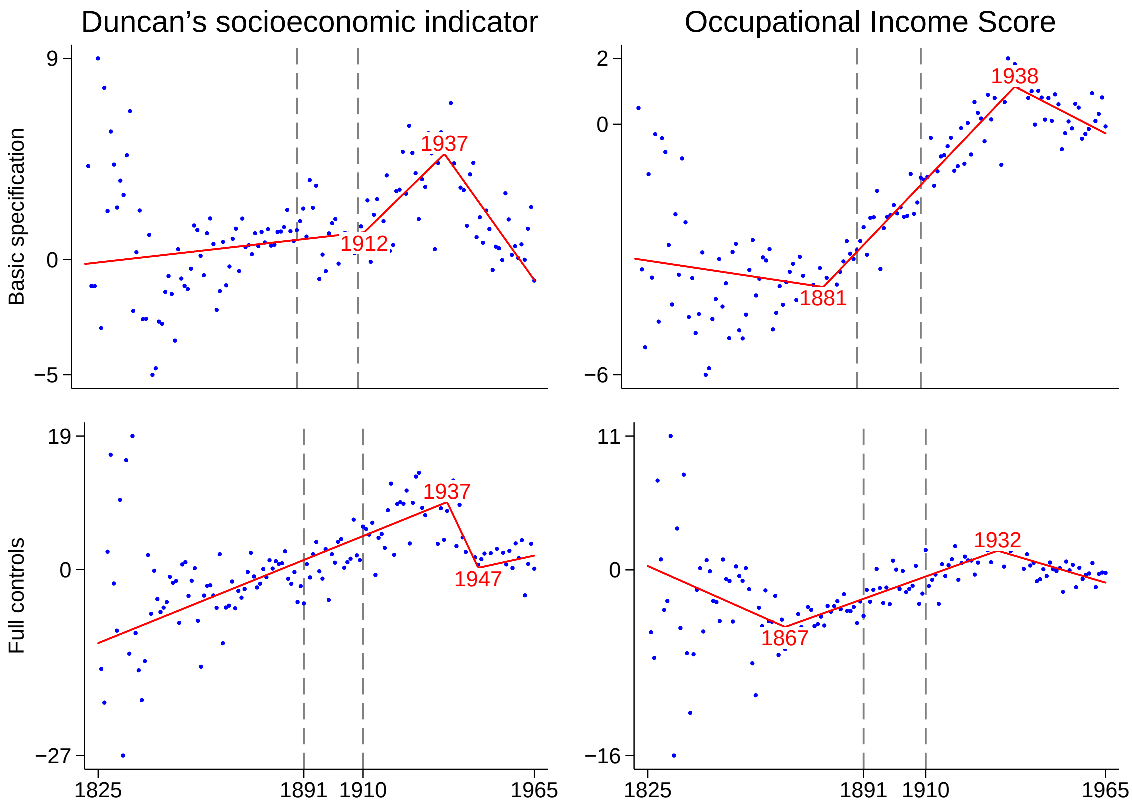
Notes: Blue dots depict same point estimates as in Figure 4. Each contour depicts the best fit of a linear model with the Bleakley (2007) exposure function, Exp, and polynomial time controls ranging in order from 0 to 5. Fits for orders 0–5 are drawn in orange, green, blue, red, purple, and brown, respectively. p-values are for the coefficient on Exp in the order-0 through order-5 models, respectively. They are based on standard errors clustered by birth state.

Figure 6: Replication and extension of Bleakley (2007) Figure III: Model with linear spline generalization of step function, fixed kink dates



Notes: Blue dots depict same point estimates as in Figure 4. Red contours depict best fits of a piecewise-linear model allowed to kink at the same dates as the Bleakley (2007) exposure function, Exp, 1891 and 1910. Each segment spans 19 years. p-values in each pane are, respectively, for the nulls of no slope change between the first segment and the second, and between the second and the third. p-values based on standard errors clustered by birth state.

Figure 7: Replication and extension of Bleakley (2007) Figure III: Model with linear spline generalization of step function, flexible kink dates, expanded data set



Notes: Blue dots depict same point estimates as in Figure 4. Red contours depict best fits of a piecewise-linear model allowed to kink twice, and fit using the mean-squared-error criterion.

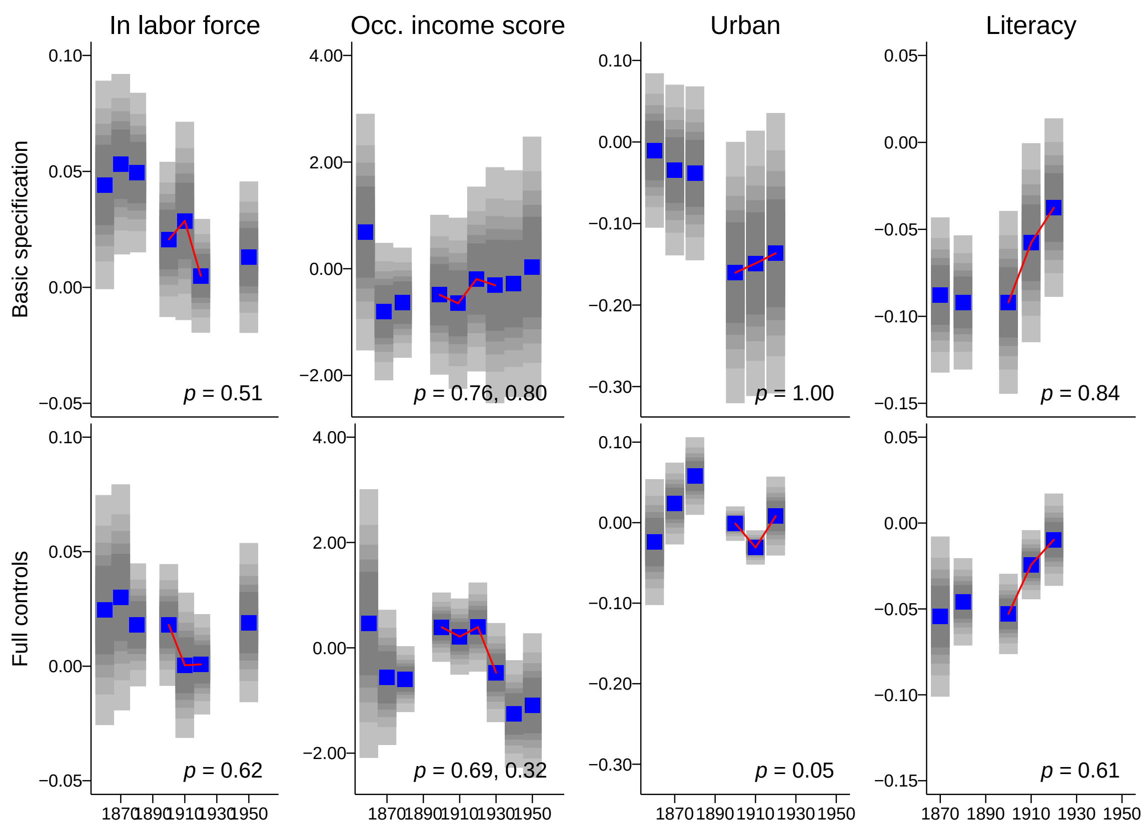
8 Appendix A - Additional Tables and Figures

Table A1: Revision of Bleakley (2007) Table VI: Exposure to RSC versus alternative time-series relationships, excluding blacks and women

Outcome	Controls	Coefficient on $H \times Exp$					
Duncan's socioeconomic indicator BIC	Basic	0.3646* (0.2045) 1385.00	0.3530 (0.2324) 1389.39	0.3286 (0.2251) 1391.38	–0.1678 (0.1728) 1342.87	**–0.1538** **(0.1651)** **1341.23**	–0.3284** (0.1632) 1349.50
Duncan's socioeconomic indicator BIC	Full controls	0.5260** (0.2220) 1588.38	0.8069*** (0.1825) 1595.30	0.6474*** (0.1845) 1589.27	**0.1478** **(0.2076)** **1578.62**	0.1168 (0.2048) 1585.24	–0.5130* (0.2767) 1589.05
Occupational Income Score BIC	Basic	0.2979* (0.1681) 1212.77	0.1371 (0.1307) 1183.66	0.1498 (0.1390) 1180.98	–0.0847 (0.0854) 1141.60	**–0.0759** **(0.0899)** **1135.39**	–0.0824 (0.0851) 1140.95
Occupational Income Score BIC	Full controls	0.2978*** (0.1067) 1366.89	0.2264** (0.0871) 1373.83	0.2445** (0.1015) 1373.91	**–0.0546** **(0.0927)** **1354.11**	–0.0530 (0.0972) 1358.83	–0.2473** (0.1025) 1358.41
Order of Polynomial Trend:		0	1	2	3	4	5

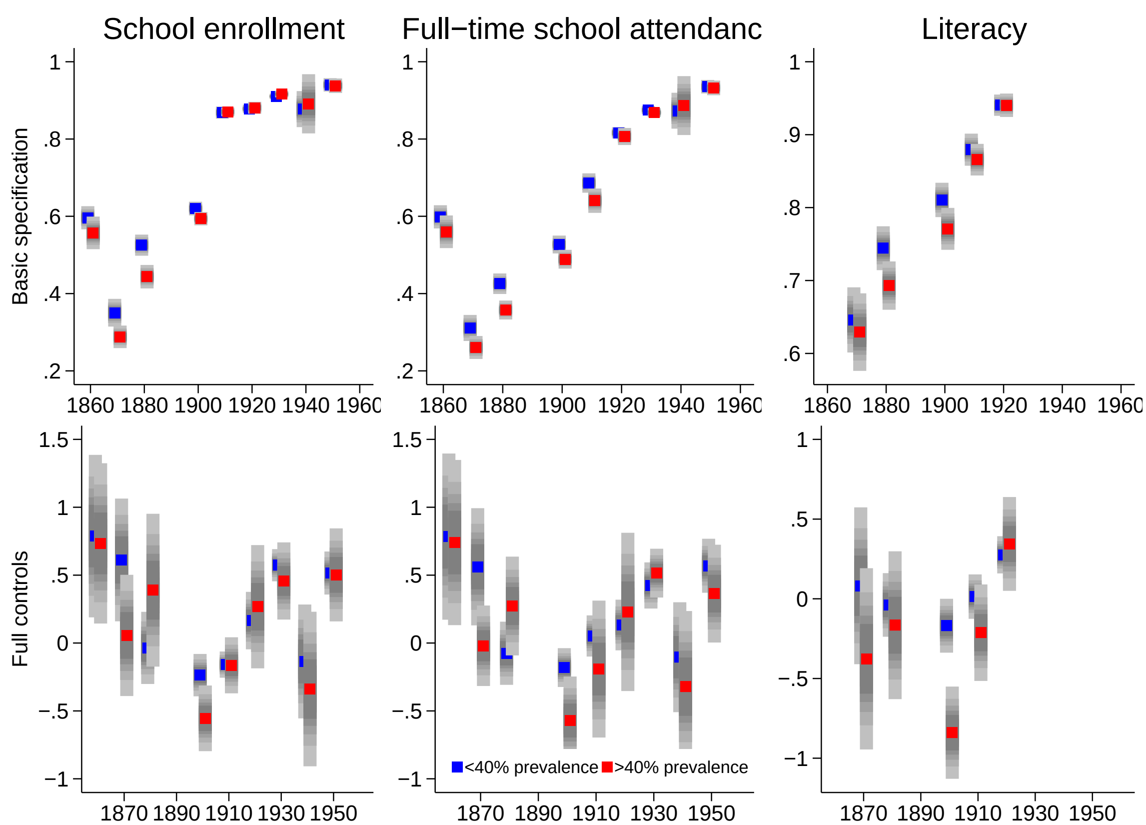
Estimates based on expanded data set, including blacks as well as whites. Regressions weighted by IPUMS-provided sampling weights. Standard errors clustered by state of birth in parentheses. BIC is the Bayesian Information Criterion, taking sample size as 141 and mean-squared error from data points and model fits presented in Figure 5. Bolded results in each row are those favored by the BIC. $^{*}p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

Figure A1: Extension of Bleakley (2007) Figure II to adult outcomes: expanded data



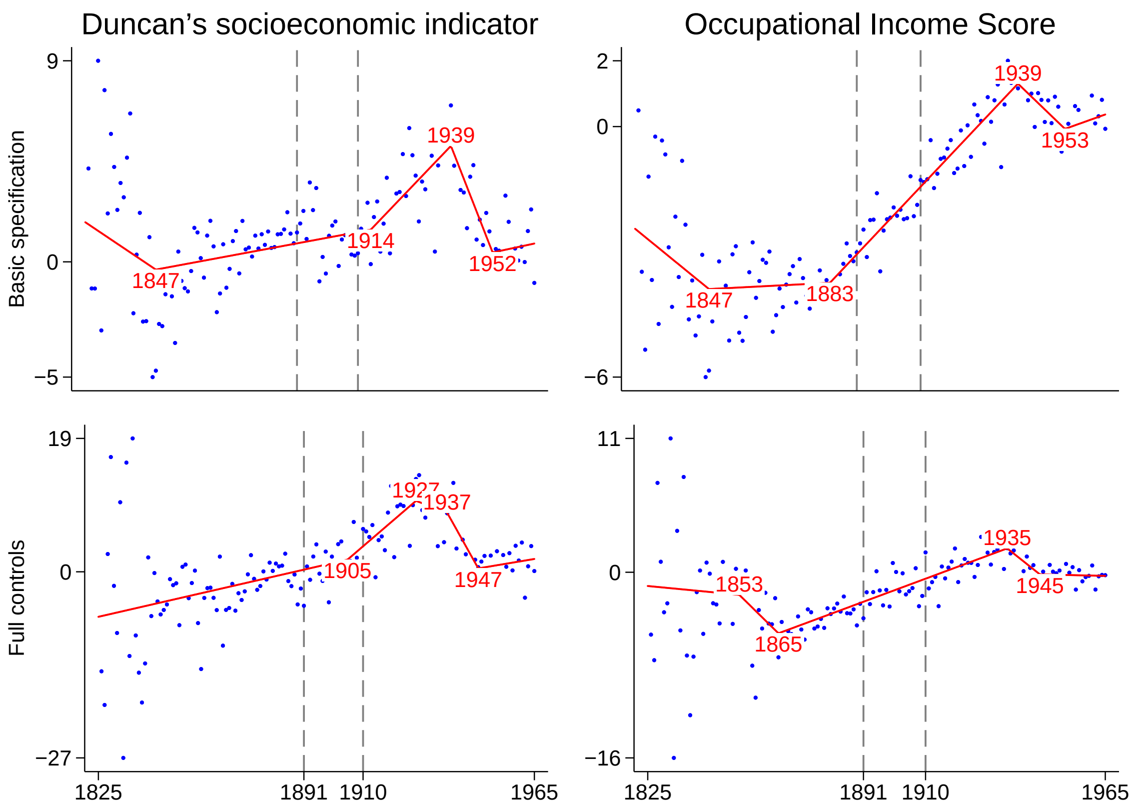
Notes: Blue dots depict point estimates of the cross-SEA association of baseline hookworm prevalence with the outcome shown within census rounds. Red lines highlight the quantities of particular interest, the rates of change just before, during, and just after 1910–20. p-values in each pane are, respectively, for the nulls of no slope change between 1900–10 and 1910–20, and between 1910–20 and 1920–30. Lack of literacy data after 1920 prevents the second test.

Figure A2: Time series variant of Bleakley (2007) Figure II, with separate regressions for below- and above-40%-prevalence samples, expanded data set



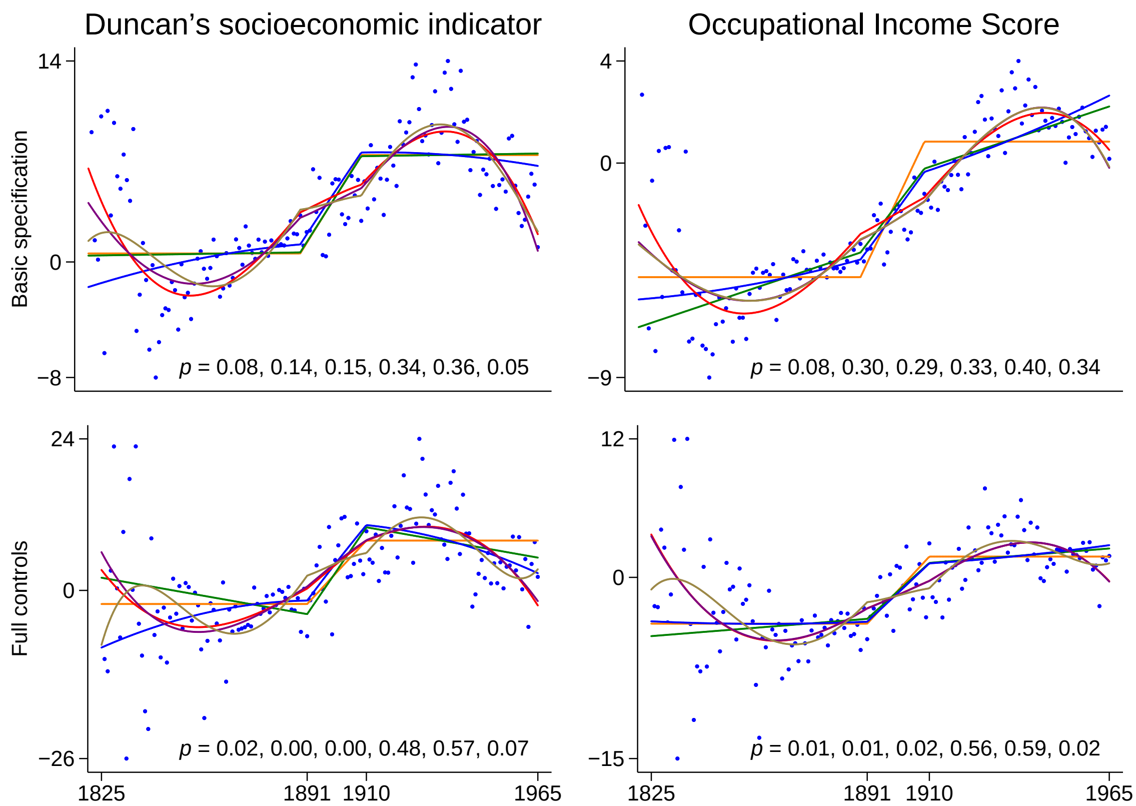
Notes: Blue dots depict point estimates of the cross-SEA association of baseline hookworm prevalence with the outcome shown within census rounds. Red lines highlight the quantities of particular interest, the rates of change just before, during, and just after 1910–20. p-values in each pane are, respectively, for the nulls of no slope change between 1900–10 and 1910–20, and between 1910–20 and 1920–30.

Figure A3: Replication and extension of Bleakley (2007) Figure III: Model with linear spline generalization of step function, four flexible kink dates, expanded data set



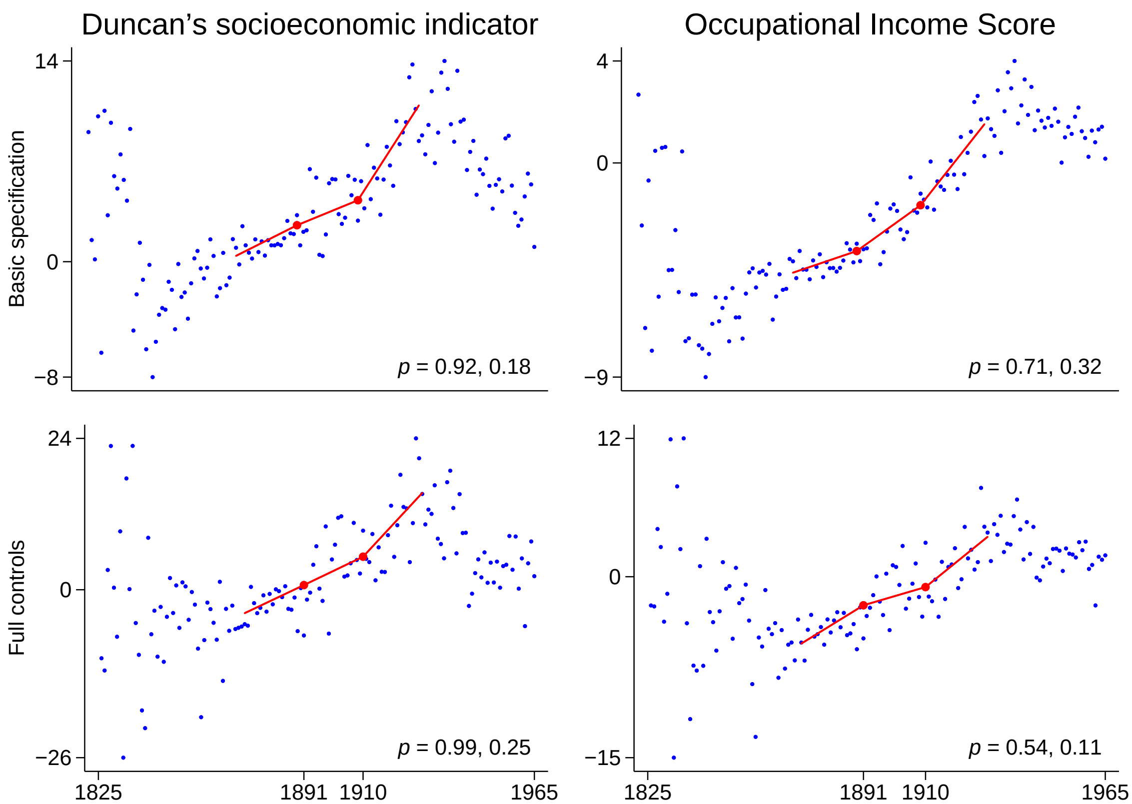
Notes: Blue dots depict same point estimates as in Figure 4. Red contours depict best fits of a piecewise-linear model allowed to kink four times, and fit using the mean-squared-error criterion.

Figure A4: Replication and extension of Bleakley (2007) Figure III: Model with polynomial time controls, fit to expanded data set, excluding blacks and women



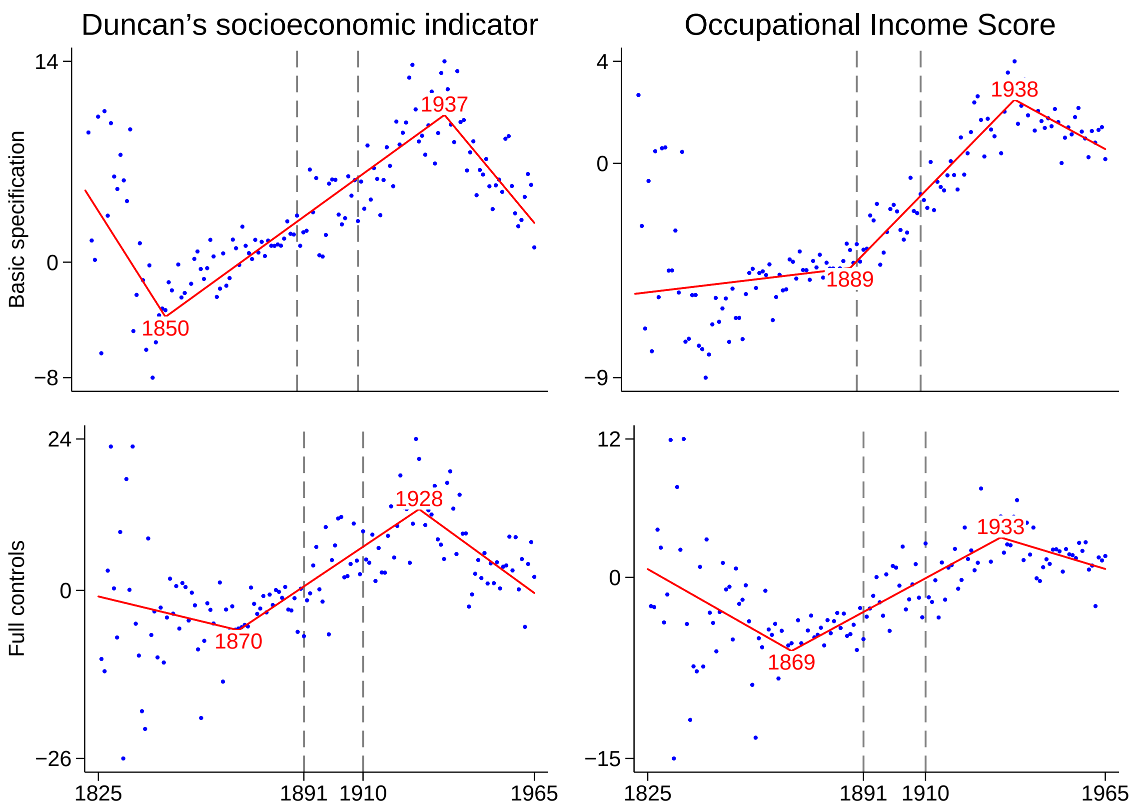
Notes: Blue dots depict same point estimates analogous to those in Figure 4, but in a sample excluding blacks and women. Each contour depicts the best fit of a linear model with the Bleakley (2007) exposure function, Exp, and polynomial time controls ranging in order from 0 to 5. p-values are for the coefficient on Exp in the order-0 through order-5 models, respectively. p-values based on standard errors clustered by birth state.

Figure A5: Replication and extension of Bleakley (2007) Figure III: Model with linear spline generalization of step function, fixed kink dates, excluding blacks and women, expanded data set



Notes: Blue dots depict same point estimates analogous to those in Figure 4, but in a sample excluding blacks and women. Red contours depict best fits of a piecewise-linear model allowed to kink at the same dates as the Bleakley (2007) exposure function, Exp, 1891 and 1910. Each segment spans 19 years. p-values in each pane are, respectively, for the nulls of no slope change between the first segment and the second, and between the second and the third. p-values based on standard errors clustered by birth state.

Figure A6: Replication and extension of Bleakley (2007) Figure III: Model with linear spline generalization of step function, flexible kink dates, excluding blacks and women, expanded data set



Notes: Blue dots depict same point estimates analogous to those in Figure 4, but in a sample excluding blacks and women. Red contours depict best fits of a piecewise-linear model allowed to kink twice, and fit using the mean-squared-error criterion.