REPLICATION AND REANALYSIS OF BLEAKLEY (2007): THE IMPACTS OF HOOKWORM ERADICATION IN THE AMERICAN SOUTH

David Roodman[1]

Open Philanthropy Project

November 21, 2017

[1] I thank Caitlin McGugan for comments on an earlier draft, Christian Smith and Claire Wang for help with data collection, and Zachary Tausanovitch for help with data collection, database construction, and analysis.

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

The *benefits* of deworming are harder to gauge. The evidence on the short-term impacts on nutrition and cognition has been rich enough to support large meta-analyses (Taylor-Robinson et al. 2015; Welch et al. 2016; Croke et al. 2016). Longer-term impacts could be much larger. But evidence on them is much scarcer. Following up on the Miguel and Kremer (2004) experiment in western Kenya, Baird et al. (2016) find impacts on earnings ten years on sufficient to generate an internal rate of return to deworming of at least 32% per annum. Ozier (forthcoming) follows up on the same experiment at about the same time, and reports cognitive gains among children who were too young to have 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 10-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 principle sources of variation: geographic variation in the initial prevalence of hookworm—thus the scope for gain from eradication—and the sudden onset of the campaign. The Bleakley (2007) results parallel those from Kenya (Miguel and Kremer 2004; Baird et al. 2016): mass deworming of children boosts schooling in the short run and earnings in the long run.

This 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.[2] Moving from replication to reanalysis, the paper then modifies specifications in order to test robustness. In particular, it:

1. *Takes advantage of the larger historical census samples now available from the Integrated Public Use Microdata Series* (IPUMS; Ruggles et al. 2015). The new data set includes records from the 1930 census, which were not available to Bleakley.
2. *Copies specification choices among the displays.* 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 the other two.
3. *Performs tests to focus more sharply on whether trends break at times explicable by the eradication campaign.*
4. *Corrects for a few econometric issues.* Most notably, a three-stage estimation process in the original is revised so that uncertainty from the initial stages is reflected in the final stage's standard errors.

A pre-analysis plan, registered with the Center for Open Science, envisioned some of these steps—the fourth and part of the third. As that statement implies, I did not limit myself to the pre-analysis plan. But the plan does credibly disclose which steps I chose before encountering the data.

The new analysis focuses on the figures more than the tables. The figures bring out temporal patterns clearly and motivate formal tests for whether the outcomes evolved in ways partly explicable by the timing of the

[1] GiveWell, "Deworm the World Initiative, Led by Evidence Action," November 2016, givewell.org/charities/deworm-world-initiative.

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

hookworm eradication campaign. And the timing—as distinct from the geographic pattern of initial hookworm burden—is the most plausibly exogenous component of the identifying variation.

The Bleakley (2007) designs are of two major forms. In “sequential cross section” (SCS) specifications, census microdata are grouped by the census year in which the data were collected. This facilitates checking, for example, whether school enrollment rose relatively rapidly between 1910 and 1920 in initially high-prevalence areas. The “retrospective cohort” (RC) specifications instead group individuals by year of birth. The first set of RC regressions use data from a single census to observe, for example, whether a historical earnings gap between low- and high-prevalence areas narrowed for people born late enough to benefit from the eradication campaign during childhood. The second set of RC regressions pools data from multiple censuses, aggregating across all census rounds that people of a given birth cohort appear in. This allows a linkage between a person’s exposure to the campaign during childhood and earnings throughout adulthood.

The new analysis recognizably matches the original’s tabulated and graphed results, but for two noteworthy exceptions. Unlike in the original, including Bleakley’s seven health and health policy controls substantially reduces the SCS impact estimates on schooling. Including *all* the Bleakley (2007) controls essentially erases the enrollment results. And an SCS finding highlighted graphically in the original—a one-time jump in school enrollment between 1910 and 1920, when no extra controls are added—is discernible in the replication, but not with compelling statistical significance.

But, moving from replication to reanalysis, the paper presents new results that strongly question the Bleakley (2007) conclusion that hookworm eradication brought detectable short- and long-term benefits. As a first step, I replicate that SCS graph for the other Bleakley (2007) indicators of human capital investment, full-time school attendance and literacy. And I test for robustness to using newer, larger census microdata samples. These steps leave the original graphical result SCS looking fragile. In particular, expanding the census samples by a factor of about 100 makes clear that the relative upward trend in schooling in historically high-burden areas began before the eradication campaign.

As for RC estimates of 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—converged over time across the gradient of historical hookworm burden. But when I formally test whether convergence temporarily accelerated as the effects of the eradication campaign set in, I do not find convincing evidence in favor. Again, the convergence began earlier than Bleakley’s impact theory predicts and continues longer.

I began corresponding with Hoyt Bleakley about these findings in May 2017. Bleakley stated that the original data and code were hard to access, and did not provide any other information that could help explain imperfections in this replication. Pending confrontation with the original data and code, I believe the reconstructed data and code introduced here ought to be viewed as the reference implementation of Bleakley (2007) since only they will be freely accessible.

Section 1 of this paper details the Bleakley (2007) designs. Section 2 explores several cross-cutting themes in the reanalysis. Sections 3 and 4 replicate and reanalyze the SCS and RC regressions. Section 5 concludes.

1 The Bleakley (2007) designs

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 many controls relating to health, education, race, and agriculture. All come from sources published about a century ago.
- A pure time series indicator for exposure to the eradication campaign (Exp), which is interacted with H to form treatment.
- 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.

The three kinds of variables are integrated at the resolution of the census data, with one observation per sampled individual. In one case, discussed below, the data are aggregated before analysis, within birth state–birth year–census year cells. In the rest, the microdata are not aggregated before regression, but standard errors are clustered after, by geographic unit and time period.

The core estimating equation can be written

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

for outcome Y for individual i in geographic unit j at time t. β is the 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 $\mathbf{z}_i$ are individual-level demographic traits, such as age, sex, race, and interactions thereof. 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 regressions 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:

$$Y_{ijt} = H_j\beta_t + \mathbf{z}_i'\boldsymbol{\alpha}_t + \mathbf{x}_j'\boldsymbol{\gamma}_t + \epsilon_{ijt} \qquad (2)$$

where the $\mathbf{x}_j$ are optional area-level 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 performs informal and formal inference about whether the β_t series constitutes evidence of impact, e.g., if it jumps around the time of the eradication campaign. The set of regressions (2) can also be performed as a single, full-sample regression in which the time dummies δ_t are interacted with all the right-side variables.[3]

The study's two designs, successive cross-section and retrospective cohort, 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.

The SCS design categorizes an observation by when and where it was collected, meaning the census year and the person's place of residence. Campaign exposure (Exp) is then simply a dummy for post-campaign censuses, i.e., for $t \geq 1920$. As for H, the census records place of residence with high precision—though the public, digitized microdata somewhat 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 surveys. That

[3] In abandoning any assumption about the functional form of Exp, estimating (2) sacrifices the ability to control for area effects δ_j. One cannot estimate area effects in a cross-area regression.

is, H could be defined by county of residence.[4] 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 data specify residence by SEA but not county. Thus, in the SCS design, j indexes SEAs. Since the RSC waged its campaign across 11 southern states, from Virginia to Texas, it surveyed prevalence only in those states. That restricts the samples of the SCS regressions.

In the retrospective cohort (RC) design, j and t index place and time of birth instead of place and time of survey.

The RC design 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, laid bricks in Atlanta in 1940, worked as a general contractor in Lexington in 1950, and ran a construction company in Phoenix in 1960, all three resulting census observations would be associated with Georgia 1915. In this way, Bleakley (2007) limits attrition from internal migration. Bleakley's single-census RC specifications use data from the 1920 or 1940 census. The multi-census specifications use all census data from 1870 to 1990 that were available to Bleakley during analysis.

The redefinition of the time and place indexes for the RC framework triggers several changes in implementation. Partly because the cadence of t shortens from decadal to annual, Bleakley incorporates more timing information into Exp. Instead of being a post-eradication census 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, $Exp = 1$ for a person born in 1892, since the person would have been 18 in 1910 and thus only enjoyed that one childhood year of notional eradication. And $Exp = 19$ for all people born in or after 1910. This construction makes Exp a piecewise-linear function of birth year, which I will call the step function. It assumes that exposure at each year of childhood matters equally for long-term outcomes. That is a reasonable choice in the face of uncertainty. But it might be substantially incorrect, as some evidence suggests that health in early childhood matters most for adult outcomes such as schooling, income, and assets (Victora et al. 2008).

The census observes place of birth, unlike place of residence, only at the state level. So in the RC regressions, the state replaces the SEA as the geographic unit. As a result, to perform the RC regressions, Bleakley widens the geographic scope to the continental United States and discards the county-level Rockefeller prevalence data as the basis for H in favor of a state-level indicator of hookworm prevalence (Kofoid and Tucker 1921). Loosely speaking, where SCS compares clusters of counties within Mississippi, RC compares Mississippi to Michigan.

2 Themes in the replication and reanalysis

2.1 Pre-analysis plan

A pre-analysis plan was registered with the Center for Open Science. It does not confine the analysis. But it credibly discloses which parts were pre-conceived and which were chosen after encountering the data.[5] 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.

---

[4] The RSC subdivides a few counties, for reporting purposes.

[5] See osf.io/yb537.

- "Performing two-stage least squares instead of the original's indirect least squares in order to obtain proper confidence intervals for instrumental variables point estimates." This step was ill-conceived. The original uses ILS in situations where conventional IV estimation is impractical, because, e.g., the impacts of the instruments on the treatment and on the outcome are estimated in different contexts.
- "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 3.2 below). Since the (temporal) variation in Exp is more plausibly exogenous than the (spatial) 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, and because even with SEA-clustered standard errors, the reanalysis casts substantial doubt on the original. However, the reanalysis of the multi-census 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 4.2. (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.

2.2 Expanded IPUMS samples

The coverage of IPUMS has expanded steadily over the years, both in the census 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 multi-census RC. Bleakley (2007) largely does not specify the densities of the samples used, but they can be estimated by reviewing the history of 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.[6]

In addition to reconstructing the original data set according to these estimates, I test robustness by switching to an expanded data set with newer IPUMS samples. (See column 2 of Table 1.) Data are added for 1860, 1930, and 2000—though the 1860 and 2000 data figure only in the multi-census RC regressions. Density rises to 5% in 1900 and 1960, and to 100% for 1910–40, using preliminary releases for the latter. For concision, I only report—and have only performed—this expanded-data test for the figures, not the tables. (Section 2.4 explains the focus on the figures.)

While the data expansion was not pre-registered, it was to a degree inevitable because the modern IPUMS interface tends to hide two samples that Bleakley (2007) appears to use: the 1-in-760 sample for 1900 and 1-in-250 sample for 1910.[7] Especially since the density of the original samples is not fully documented, a contemporary user naturally gravitates to some of the newer, larger samples.

---

[6] I am least sure about the 1900 sample used in the SCS regressions. Bleakley (2007) may have used the preliminary version of the 1-in-100 IPUMS 1900 sample, which is a 1-in-200 sample that was posted on May 7, 2002 (usa.ipums.org/usa-action/revisions). However, I achieve better matches with the older, 1-in-760 sample, and so use that.

[7] As of October 2, 2017, 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—but not the 1% sample—by the field SAMP1910.

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. However, there are exceptions (Ruggles et al. 2015; usa.ipums.org/usa/intro.shtml#weights). And since different censuses are sampled at different densities, pooling them effectively introduces sampling imbalances, which can be corrected by using the IPUMS-provided weights. Bleakley (2007) does not mention using sampling weights.

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%
1920	1%	100%
1930	0%	100%
1940	1%	100%
1950	1%	1%
1960	1%	5%
1970	1%	1%
1980	5%	5%
1990	5%	5%
2000	0%	5%

[1] Excludes slaves.

[2] Pairs of numbers refer separately to SCS and RC regressions.

2.3 Differences among specifications

Nearly all the Bleakley (2007) results appear in tables and figures. Naturally, the specification differ from each other in various respects. Some of these differences are dictated by the empirical questions that motivate individual regression, or by the structure of the data. Others, however, are more discretionary. The discretionary variety in the original generates some minimally discretionary robustness testing in the reanalysis. Distinctive choices in one specification can be copied to others.

The Bleakley (2007) figures and tables are listed here in Table 2. Perusing the table exposes these within-study differences:

- Using SCS regressions, Bleakley (2007)'s Tables II and III examine impacts on three outcomes: school enrollment, full-time school attendance, and literacy. The parallel Figure II looks only at the first.
- Similarly, Table III carries out "full controls" SCS regressions, which, as one would expect, add many additional controls. The parallel Figure II does not.
- The situation is the other way around for the RC regressions: the figure (Figure III) includes full-control results while the table (Table V) does not.
- Only Figure III is restricted to whites.[8]

[8] Bleakley (2007) does not motivate the exclusion of blacks. However, Bleakley (2010, note 7), referring to an application of the same design to malaria eradication, writes:

- Most of the displays show results from regressions on individual-level data. Figure III is the exception: before its regressions are run, the data are aggregated by birth year and birth state. This sacrifices the ability to control for individual demographics.[9]

Below, I carry out regressions that bring symmetry to the specifications reported in the original. I generate Figure II for all outcomes, without and with full controls. I replicate Table V with full controls. And I run Figure III at the individual level, while also adding blacks.

None of these robustness tests is pre-registered.

> 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.

Since both malaria and hookworm were concentrated in the South, the same arguments may well have motivated the restriction to whites in Figure III of Bleakley (2007). In section 4.2, I include blacks as I expand the IPUMS samples for the multi-census RC regressions. The reasons are two. First, the IPUMS USA census data do not include slaves. A separate IPUMS project has digitized census records of slaves, but unfortunately the census enumerators gathered quite different information about slaves, so the data sets do not harmonize. So Bleakley's concern about comparing slaves to non-slaves does not apply. Of course, it may be also problematic to compare free blacks in slave and non-slave states, but that is just one more example of how the regions of the United States are poor comparators for each other, and highlights the importance the temporal dimension of identification in Bleakley. Second, Bleakley (2007) Figure II also includes blacks and, in a slightly revised edition (Bleakley 2009), reaches back to 1870. It provides precedent for doing the same in replicating Figure III.

[9] These regressions are restricted to whites, so race is moot as a control. And as section 4.2 explains, age is controlled for in an unconventional way, being partialled out of the dependent variables before aggregation. This leaves sex as a natural demographic control to include if regressing on microdata.

TABLE 2. DISPLAYS IN BLEAKLEY (2007)

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

2.4 What constitutes evidence of impact in a time series?

A deep question that must be answered in order to interpret the Bleakley (2007) results is: what constitutes persuasive evidence that an impulse of a given functional form is a component of an observed time series?

I will motivate this abstract question with a hypothetical example. Suppose the time series for our outcome of interest follows a logistic curve over the years 1825–1965, as in Bleakley (2007) Figure III. It can obey the logistical form perfectly, or with an independent, normal error component. To test whether events circa 1910 can help statistically explain such a series, I fit two models with OLS:

- Model 1 defines the independent variable of interest—treatment—as a post-1910 dummy, and controls with constant and linear trend terms.
- Model 2 defines treatment as the step function described in section 1, the piecewise linear form that is allowed to kink in 1891 and 1910. The model controls with a constant.

Figure 1 plots the two variants of the outcome—with and without a stochastic component—in blue. The red and black curves indicate the best fits of the two models.[10] Table 3 shows the corresponding regression estimates for treatment impact, using non-robust standard error estimates. In all, the term of interest is highly significant. If one ran these regressions *without* graphing the data and the best fits, one could easily interpret the results as evidence that both forms of the treatment variable are good explanators for the history of the outcome. But inspection of the graph makes obvious that both models are misspecified. No historically unusual trend breaks occurred at the times implied by the models.

[10] The step function is normalized to rise from 0 to 1. The outcome is generated as $1/(1 + \exp(-(t - 1870)/10))$. The optional normal error has standard deviation 0.2.

The example is fanciful. Yet Model 1 is mathematically analogous to the more conservative of the Bleakley (2007) SCS specifications (equation 2 in the original). There, the time series is the cross-sectional association between baseline prevalence and outcomes of interest in successive censuses, climbs in which indicate relative gains for historically high-prevalence areas. And in the time dimension, the treatment variable is a post-1910 dummy. Geographic units are allowed individual time trends. Meanwhile, Model 2 is mathematically analogous to the RC specification (equation 3 in the original), in which the temporal dimension of treatment is modeled as a step function that rises between 1891 and 1910.

Thus while the numerical results in Table 3 can be correctly read to say that the outcome rose when the impact models predict that it will, the results do not credibly demonstrate *causality*. This is why, in revisiting Bleakley (2007), I focus on the time series plots, and the question of how to infer impact from them. (I reconstruct all the tables too, but mainly to check on the quality of the replication.)

Accepting that we should center our inference on the time series, however, does not tell us *how* to do it. How should we judge whether a treatment impulse contributes substantially to an outcome time series, in which other influences are also at play? The issue can be seen as having two aspects, one technical, and one that I will call "Bayesian." As a technical matter, to test hypotheses about impact, it seems that we must restrict ourselves *a priori* to certain parametric families of models, such as the two above.[11] We can then test whether, within such a family, members that correspond to zero impact are rejected by the data—i.e., significantly reduce the quality of the model fit.

In reanalyzing Bleakley (2007), I work within two parametric families of models. One, not pre-registered, generalizes the step function to a piecewise-linear spline just like that in Model 1. Kinks are allowed at the times chosen, 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, just as in Model 2. I test whether the slope rises at the first kink and falls at the second, as it should under a step-like impact contour. And I plot the model fits. The second parametric family does not generalize the functional form for treatment, and instead introduces controls for polynomials in time up to order five. This approach is implicitly pre-registered in that Bleakley also employs it.[12] Neither approach is obviously optimal. One could construct others, such as controlling for a logistic function of time. But these two choices seem minimally discretionary, intuitive, and informative.

The Bayesian aspect of this analytical challenge is that how prepared we are to take such test results as evidence of impact should depend on how much credence we place on alternative explanations. In the hypothetical example in Figure 1, if we are nearly certain that no other theory can explain the century-scale rise in the outcome variable, that should increase our readiness to believe that the treatment of interest was the causal factor, however misspecified it may seem. On the other hand, if competing theories are in the offing, that should increase our demand for a good match on functional form.

I see the case of hookworm eradication in the American South as closer to the latter extreme. Within the South, low- and high-hookworm areas differed systematically in geography. 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)

---

[11] The fitting process could involve a nonparametric step, e.g., to filter out dynamics in certain frequency ranges. But inference would then still need to be performed with respect to a more limited set of parameters.

[12] Bleakley (2007) only reports specifications up to order three, but footnote 25 reports testing up to order five.

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. Yet it would also not be surprising if the economic importance of these historical differences dwindled in the twentieth century, as agriculture's share in the economy shrank. In my view, then, a finding of long-term convergence between historically low- and high- hookworm areas does not add much credibility to the proposition that eradication brought large economic benefits.

Perhaps for this reason, Bleakley (2007) tests most of its regressions for robustness to controlling for initial conditions. While welcome, such controls may not suffice. For the SCS regression, which identify off of SEA-level variation, the convergence controls are all observed at the state level, so they cannot adjust for within-state convergence. The RC regressions take the state as the geographic unit, and control for convergence with dummies for the major U.S. regions and a state-level measure of agricultural wages in 1899. The latter is observed at the same geographic unit as treatment, so it might control for convergence at the desired level of granularity. But a measure of *pay* in *one* occupation in 1899 may not fully predict convergence in occupational *standing* over the following half century—and occupational standing, rather than earnings, is what is tracked over the long term.

On balance, the Bleakley (2007) controls cannot banish the concern that forces outside the analysis drove convergence. As a result, evidence of convergence alone would not persuade me as much as would evidence of *acceleration* in convergence with timing that fits the onset of the hookworm eradication campaign in the early 1910s.

FIGURE 1. BEST FITS OF TWO MODELS TO A HYPOTHETICAL OUTCOME WITHOUT AND WITH A RANDOM COMPONENT

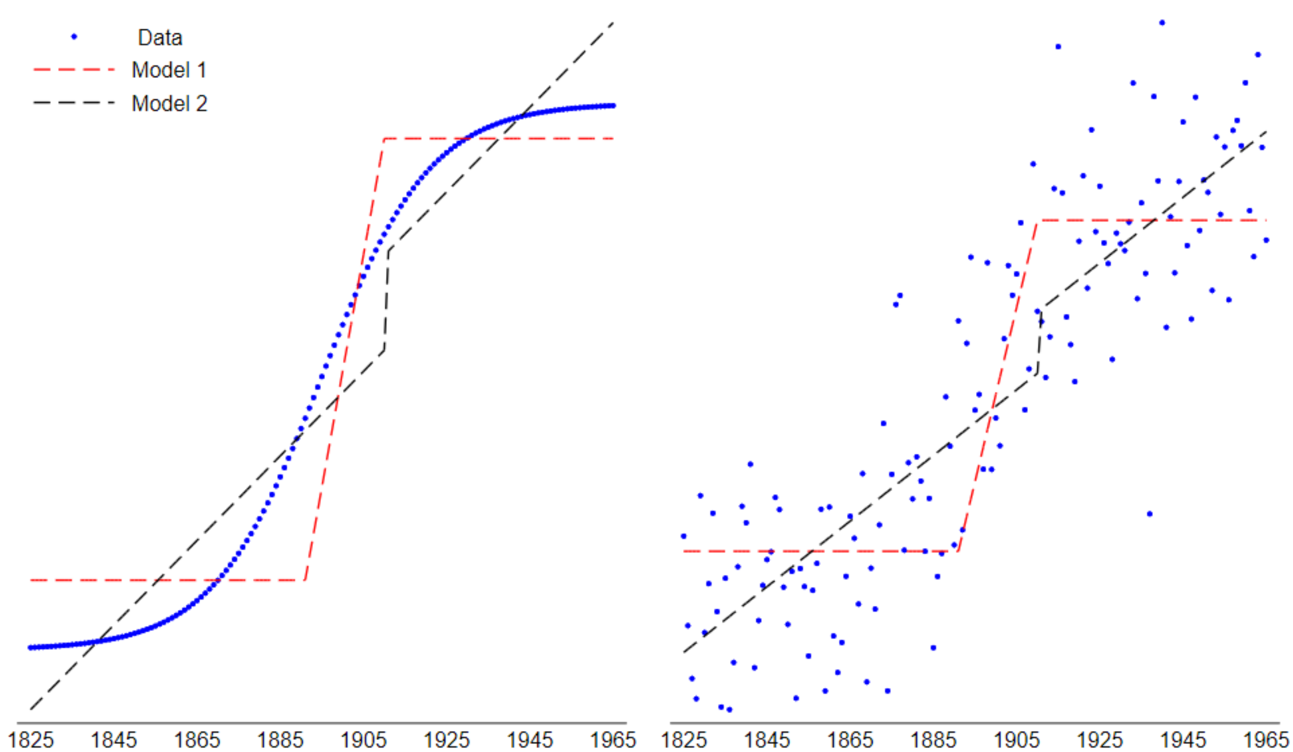


TABLE 3. IMPACT ESTIMATES FROM TWO MODELS APPLIED TO A HYPOTHETICAL OUTCOME WITHOUT AND WITH RANDOM COMPONENT IN THE DATA GENERATING PROCESS (DGP)

	Deterministic DGP		Stochastic DGP	
	Model 1	Model 2	Model 1	Model 2
Treatment	0.807***	0.174***	0.780***	0.144**
	(0.0189)	(0.0280)	(0.0399)	(0.0699)
Observations	141	141	141	141

Classical standard errors in parentheses. ** $p < 0.05$, *** $p < 0.01$.

3 Replication and reanalysis: Successive cross-section specifications

Recall that the successive cross-section (SCS) analysis 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. The regressions are confined to 11 states in the American South. The exposure variable Exp is a dummy for censuses fielded after the eradication campaign.

In reconstructing the data set from the sources listed in Bleakley (2007), my assistants and I encountered ambiguities in the definitions of some variables and missingness in the sources. As a result, it is impossible to exactly reproduce the original data set without access to it. For example, my assistants and I could not find data for Mississippi for one of the education controls, the value of school plant and equipment. Likewise for Kentucky and school term length. And 1902 health spending data is zero or missing for almost all Arkansas counties, which causes Bleakley's control, the log change in per-capita health spending between circa 1902 and circa 1932, to be missing too. To document our choices, we publicly post a spreadsheet that forges our replication variables from the primary data and links to scanned copies of the sources.

Bleakley (2007) does not mention dealing with missingness. Here, I perform casewise deletion: in a given regression, SEAs missing values for any controls are dropped.

3.1 Replicating Tables I–IV: Short-term impacts on children and adults

Table 4, below, follows the format of Bleakley (2007) Table I in order to compare the original and reconstructed data sets on first and second moments of several variables. The table contains three pairs of columns, the first for the whole sample, and the second for low- and high-prevalence subsamples. A child infection rate of 40% marks the divide between the two subsamples. 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." As explained above, the last four variables, relating to education, were hardest to reconstruct, so it is not surprising that the matches for them are also less precise. And the reconstructed data set includes two more SEAs, both of which fall into the below-40% subsample. All these discrepancies seem impossible to fully explain without access to the original data and code.

In similar fashion, Table 5 replicates the first set of Bleakley (2007) SCS results, from the original's Table II. Each cell reports the coefficient on $H_j \times Exp_t$ in a distinct regression. For the simplest specification, in the first row, the replication again matches well. Here, the sample is restricted to the censuses of 1910 and 1920, which bracketed the eradication campaign. With reference to equation (1), 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 four rows of Table 5 also present reasonable matches. The second row expands the sample to 1900–50. Viewing the specification as difference-in-differences, 1900–10 now constitutes the pre-treatment and 1920–50 the post-treatment period. Literacy is dropped as an outcome because it is not available in the census after 1920. The third row inserts a set of SEA-specific linear time controls in $\mathbf{x}$. This is the Bleakley specification most analogous to Model 1 in section 2.4 above. The next rows, in panel B, retain the SEA-specific linear time controls, except for the literacy regressions, where lack of data for 1930–50 reduces the number of time periods.[13] The first of these rows introduces state × year fixed effects to control for state-level policy shocks.[14] The penultimate row instead controls for the product of Exp and the state-level school enrollment rate, in order to mitigate mean reversion.

The last row of Table 5 makes the most radical changes to the specification, and also yields the poorest matches—for reasons, again, that cannot be determined. Here, as in the RC regressions, the sample expands to the entire country and baseline infection is measured at the state of birth rather than the SEA of residence. The mean reversion control is retained. The original and new point estimates for the impact on school enrollment clash by a factor of four, and for full-time school attendance by about 1.5. For literacy, the new estimate is a third smaller than the original.

Bleakley's Table III tests the results in Table II for robustness to additional controls and, as well, explores the results for heterogeneity. These regressions are reconstructed and reported in Table 6 below. Most of the new results broadly corroborate the originals. The greatest differences appear in Panel B, which adds more controls. In particular, adding the seven health and health policy listed in Bleakley (2007) appendix section II.A reduces the impact estimates on schooling by a typical 50% (first row of panel B). Despite the difficulties in reconstructing the education controls, the regressions introducing them mostly match the original well (second row of panel B). But the full-controls regressions, which also include the health and health policy controls, find no clear impacts (las row of panel B).

Last among the SCS tables, Table 7 below largely reproduces Bleakley (2007) Table IV, which checks for impacts on adult outcomes. Like the original, the new regressions produce no robust evidence of benefits for adults—in literacy, labor-force participation, occupational income score (OIS), or urban residence. (OIS is a proxy for income: it is the median income, in 1950, for a person's reported occupation.) As Bleakley points out, these results are consistent with the theory that hookworm eradication indeed caused the relative gains found among children in the earlier tables, since adults were infected much less, and stood to benefit much less from eradication.

By and large, the original SCS results are recognizably in these reconstructions. The major exception is that now the results for human capital investment in children appear fragile to the inclusion of the full Bleakley (2007) control set.

[13] In fact, literacy data are available for 1900 and could be added to these regressions.

[14] Bleakley (2007) labels the row "Include state × Post dummies," where Post is a dummy for census year ≥ 1920. However, the text says that "(state × year) fixed effects" are added. I take the latter to be correct, and interpret it as a set of dummies for each state-year combination.

TABLE 4. 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	0.338	0.554	0.556	0.164	0.178
	(0.230)	(0.226)	(0.137)	(0.137)	(0.117)	(0.118)
Individuals Treated At Least Once by the RSC, Per School-Age Child	0.206	0.155	0.342	0.281	0.109	0.063
	(0.205)	(0.227)	(0.199)	(0.292)	(0.147)	(0.087)
School Enrollment, 1910	0.721	0.717	0.711	0.706	0.729	0.725
	(0.104)	(0.096)	(0.099)	(0.091)	(0.108)	(0.099)
Change in School Enrollment, 1910–1920	0.089	0.126	0.103	0.143	0.078	0.114
	(0.080)	(0.070)	(0.090)	(0.076)	(0.072)	(0.064)
Full-time School Attendance, 1910	0.517	0.499	0.469	0.456	0.551	0.530
	(0.140)	(0.126)	(0.123)	(0.110)	(0.141)	(0.128)
Change in Full-time School Attendance, 1910–1920	0.203	0.238	0.246	0.275	0.172	0.211
	(0.097)	(0.086)	(0.093)	(0.085)	(0.089)	(0.076)
Literacy, 1910	0.853	0.849	0.824	0.822	0.875	0.870
	(0.104)	(0.097)	(0.101)	(0.093)	(0.102)	(0.094)
Change in Literacy, 1910–1920	0.060	0.057	0.081	0.073	0.045	0.045
	(0.067)	(0.056)	(0.075)	(0.061)	(0.057)	(0.049)
Population Black, 1910	0.357	0.348	0.410	0.414	0.318	0.300
	(0.221)	(0.209)	(0.208)	(0.185)	(0.223)	(0.213)
Fraction Population Urban, 1910	0.174	0.137	0.167	0.130	0.180	0.142
	(0.200)	(0.143)	(0.214)	(0.142)	(0.223)	(0.144)
School term, in Months, c. 1910	5.251	5.328	5.055	5.168	5.391	5.462
	(1.066)	(0.770)	(1.042)	(0.584)	(1.068)	(0.880)
School per Square Mile, c. 1910	0.195	0.146	0.142	0.132	0.233	0.157
	(0.358)	(0.048)	(0.053)	(0.037)	(0.465)	(0.052)
Value of School Property, per Pupil, Current Dollars, c. 1910	5.518	6.632	4.699	5.524	6.104	7.400
	(4.037)	(4.722)	(3.159)	(3.197)	(4.496)	(5.432)
Teacher-to-School Ratio, c. 1910	1.336	1.316	1.397	1.307	1.293	1.322
	(0.545)	(0.360)	(0.505)	(0.334)	(0.572)	(0.381)
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 5. 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***	0.0934***	0.1591***	0.1608***	0.0587***	0.0608***
		(0.0225)	(0.0260)	(0.0252)	(0.0282)	(0.0186)	(0.0195)
Observations			65436		65436		50058
1900–1950	No	0.0608**	0.0938***	0.1247***	0.1146***		
		(0.0261)	(0.0232)	(0.0286)	(0.0240)		
Observations			94665		94665		
1900–1950	Yes	0.0954***	0.1299***	0.1471***	0.1528***		
		(0.0233)	(0.0333)	(0.0287)	(0.0379)		
Observations			94665		94665		
Panel B: Effects within and between states							
Include state x Year dummies		0.1313***	0.1719***	0.2144***	0.2371***	0.0417**	0.0480**
		(0.0245)	(0.0367)	(0.0290)	(0.0370)	(0.0207)	(0.0214)
Observations			94665		94665		50058
Allow for state-specific mean reversion		0.1148***	0.1370***	0.1813***	0.1760***	0.0408**	0.0352*
		(0.0265)	(0.0368)	(0.0312)	(0.0365)	(0.0206)	(0.0203)
Observations			94665		94665		50058
Use infection from state of birth instead of SEA		0.0489	0.1826**	0.2057***	0.3066**	0.0907**	0.0630
		(0.0504)	(0.0832)	(0.0765)	(0.1212)	(0.0451)	(0.0405)
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. Equation numbers refer to Bleakley (2007). Standard errors in parentheses, clustered by state economic area, except in the last row, where they are clustered by state. Sample sizes not available from original. 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 6. 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***	0.1299***	0.0883***	0.0934***	0.1471***	0.1528***	0.1591***	0.1608***	0.0587***	0.0608***
	(0.0233)	(0.0333)	(0.0225)	(0.0260)	(0.0287)	(0.0379)	(0.0252)	(0.0282)	(0.0186)	(0.0195)
Observations		94665		65436		94665		65436		50058
Panel B: Specifications with additional controls										
Health & health policy	0.1200***	0.0490	0.1187***	0.0623**	0.1628***	0.0883*	0.1646***	0.1170***	0.0724***	0.0761***
	(0.0291)	(0.0527)	(0.0262)	(0.0276)	(0.0355)	(0.0507)	(0.0294)	(0.0335)	(0.0233)	(0.0240)
Observations		70945		49049		70945		49049		37555
Education & race	0.1235***	0.1062**	0.0793***	0.0730***	0.1851***	0.1470***	0.1581***	0.1345***	0.0556**	0.0146
	(0.0208)	(0.0435)	(0.0208)	(0.0247)	(0.0247)	(0.0424)	(0.0250)	(0.0318)	(0.0171)	(0.0230)
Observations		76516		53001		76516		53001		40534
Full controls	0.1014**	0.0315	0.0850***	–0.0631*	0.1408***	0.0815	0.1026**	–0.0237	0.0513*	–0.0429
	(0.0349)	(0.0856)	(0.0224)	(0.0342)	(0.0421)	(0.0756)	(0.0325)	(0.0450)	(0.0213)	(0.0296)
Observations		53647		37166		53647		37166		28459
Panel C: Demographic subgroups										
Preteens	0.0932***	0.1151***	0.0890***	0.0967***	0.1416***	0.1415***	0.1549***	0.1627***	0.0912***	0.0884***
	(0.0255)	(0.0381)	(0.0242)	(0.0268)	(0.0302)	(0.0421)	(0.0266)	(0.0278)	(0.0253)	(0.0262)
Observations		54653		38032		54653		38032		22654
Adolescents	0.0986***	0.1567***	0.0877**	0.0938**	0.1573***	0.1734***	0.1682***	0.1636***	0.0323	0.0389*
	(0.0280)	(0.0408)	(0.0282)	(0.0326)	(0.0336)	(0.0464)	(0.0295)	(0.0349)	(0.0165)	(0.0184)
Observations		40012		27404		40012		27404		27404
Blacks	0.2299***	0.2111***	0.1838***	0.1622***	0.2601***	0.2188***	0.2205***	0.1872***	0.1078**	0.1198**
	(0.0399)	(0.0637)	(0.0337)	(0.0390)	(0.0399)	(0.0615)	(0.0320)	(0.0372)	(0.0374)	(0.0413)
Observations		31852		22833		31852		22833		17533
Whites	0.0378	0.1018***	0.0270	0.0518	0.1103***	0.1412***	0.1169***	0.1427***	0.0264	0.0273
	(0.0237)	(0.0337)	(0.0267)	(0.0311)	(0.0294)	(0.0394)	(0.0294)	(0.0315)	(0.0139)	(0.0151)
Observations		62813		42603		62813		42603		32525

"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. Sample sizes not available from original. 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 IV: CONTEMPORANEOUS EFFECT ON ADULT OUTCOMES

	Whole		Male		Female		White		Black	
	Original	New	Original	New	Original	New	Original	New	Original	New
Literacy	0.0062	0.0207**	–0.0107	0.0122	0.0203	0.0305***	0.0107	0.0121	–0.0014	0.0447**
	(0.0095)	(0.0089)	(0.0108)	(0.0110)	(0.0127)	(0.0113)	(0.0112)	(0.0114)	(0.0229)	(0 .0193)
Observations		98562		49661		48901		65865		32697
Labor-force	–0.0069	–0.0088	–0.0069	–0.0072	–0.0056	–0.0073	–0.0212*	–0.0279**	0.0036	0.0260
participation	(0.0134)	(0.0129)	(0.0065)	(0.0064)	(0.0284)	(0.0299)	(0.0124)	(0.0122)	(0.0249)	(0.0221)
Observations		98562		49661		48901		65865		32697
Occupational	0.0526	0.2499	–0.0186	0.3543	0.0581	–0.0153	0.0855	0.6127	0.0224	–0.3008
income score	(0.2836)	(0.3822)	(0.4912)	(0.4634)	(0.4163)	(0.5521)	(0.3903)	(0.6158)	(0.3861)	(0.3341)
Observations		60947		48397		12550		37090		23857
Lives in urban	0.0157	0.0013	0.0030	–0.0033	0.0280	0.0059	0.0199	0.0019	0.0132	0.0083
area	(0.0172)	(0.0139)	(0.0190)	(0.0166)	(0.0177)	(0.0155)	(0.0226)	(0.0198)	(0.0245)	(0.0228)
Observations		98562		49661		48901		65865		32697

"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. Sample sizes not available from original. Standard errors in parentheses, clustered by state economic area. *$p < 0.1$. **$p < 0.05$.

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

Bleakley (2007) includes a single plot based on SCS regressions. It corresponds most closely to the upper left of Table II in the original and Table 5 here, and is derived by fitting (2) to data for 1900–50. The dependent variable is school enrollment, and only demographic controls are included. As foreshadowed in section 2.3, I attempt to replicate that graph and then introduce three innovations:

- Rendering it for the other SCS outcomes, in analogy with Bleakley (2007) Figure III.
- Including the full control set, also in analogy with Bleakley (2007) Figure III.
- Using the larger IPUMS sample.
- Incorporating a formal test for the step shape. (The other formal test considered in section 2.4, introducing polynomial time controls, is less useful when the data are observed at so few time points.)

In addition, to add historical perspective, I expand the graphs from 1900–50, as in Bleakley (2007), to 1870–1950, as in Bleakley (2009).[16]

The closest replication of Bleakley (2007)'s Figure II appears here in the upper-left pane of Figure 2. The blue dots are point estimates and the vertical grey bars 95% confidence intervals. Shading within the bars indicates confidence. (In the original, the confidence level of the confidence intervals is not stated.) Standard errors are clustered by census year– SEA combination.[17] Consistent with the Bleakley (2007) conclusion, 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, the acceleration in 1910 is less clear. The null hypothesis of no slope change at the 1920 kink is rejected by a two-tailed Wald test at $p = 0.03$, but only rejected at 0.37 for the 1910 kink. (Both p values are displayed in the bottom left of that pane.) On this test, it is not clear that convergence accelerated with eradication.

The rest of the first row of Figure 2 moves to the other human capital outcomes. For full-time school attendance, both slope changes are statistically significant. On the other hand, the literacy trend shows no break with the past. (Lack of data prevents checking for deceleration around 1920.)

The second row of Figure 2 adds Bleakley's full control sets. As in Table 6, panel B, this destroys most suggestion of an impact on human capital investment.

Figure 3 is constructed in the same way as Figure 2, except that it uses the expanded IPUMS samples. The 100% samples available at this writing for 1910 and 1920 lack the literacy variable, so the literacy panes on the right of the figure do not speak to the impacts of hookworm eradication, and are included only for completeness. For school enrollment and full-time attendance, the number of census observations increases from 536,005 to 58,034,919.

Perhaps the larger samples stabilize the results. More than in the previous figure, the graphs in Figure 4 tell a common story: the association between hookworm prevalence and schooling did rise between 1910 and 1920, and rose less in 1920–30—indeed, fell. But the rises begin well before the campaign.

Next, I treat in the same way the contemporaneous adult outcomes studied in Bleakley (2007) Table IV. Like that table, these figures show no clear association between the treatment variable and the outcomes.

---

[16] Bleakley (2009) was published by the World Bank and is nearly identical to Bleakley (2007).

[17] 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 × age, black, and black × age."

Last, I perform time series variants of these regressions. These do little to update priors, but are included since they were pre-registered. These 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:

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

$\mathbf{x}$ 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 very similar trends in both groups with a large jump in schooling in 1900–10, and smaller ones after. (See Figure A 2, which uses the expanded data set for precision.)

Overall, the replication and extension of Figure II substantially weakens the case that hookworm eradication quickly boosted human capital investment in children.

FIGURE 2. REPLICATION AND EXTENSION OF BLEAKLEY (2007) FIGURE II

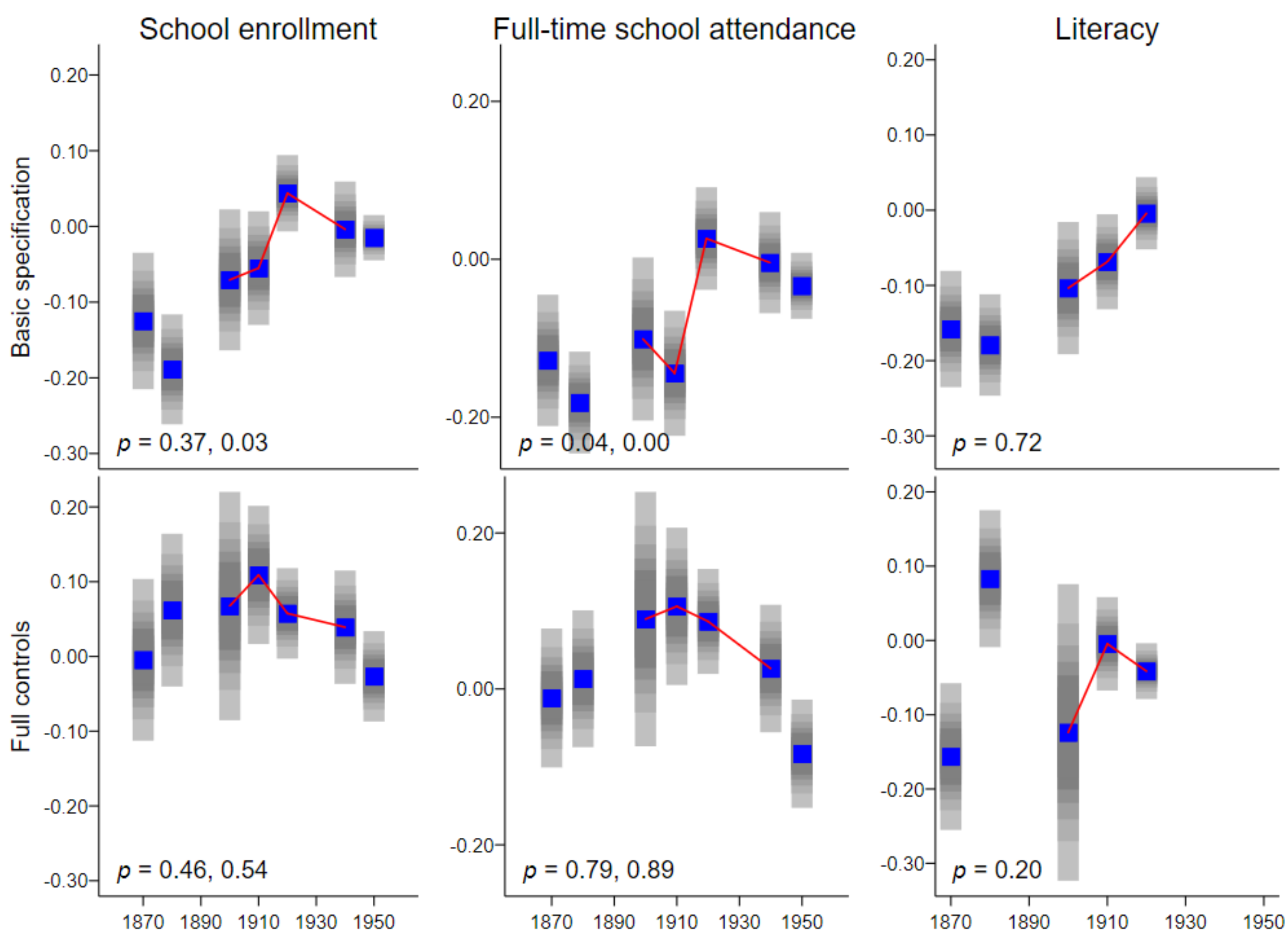
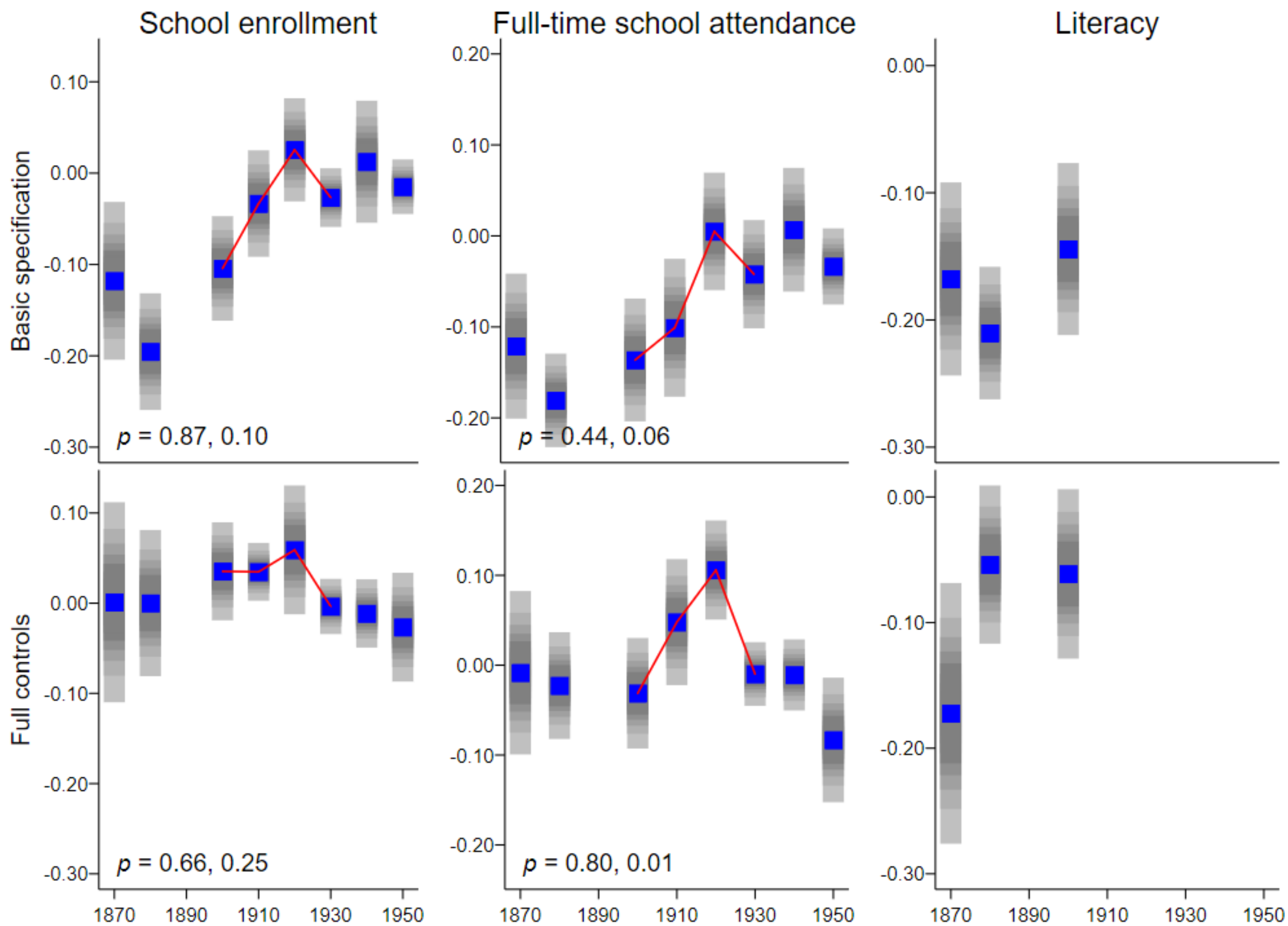


FIGURE 3. REPLICATION AND EXTENSION OF BLEAKLEY (2007) FIGURE II: EXPANDED DATA SET



4 Retrospective cohort specifications

As noted earlier, 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 regression takes data from a single census. The corresponding figure plots data from many censuses.

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, some of which comes from state government reports, which vary in coverage and format. Still, ambiguities surface here too, which again impede exact replication. Some of the education variables take data from federal reports for "circa 1902–32," so the original and reconstructed data sets may take observations from 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.[18]

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

4.1 Replicating Table V: Single-census retrospective cohorts

Bleakley (2007)'s single table of RC regression is replicated below in Table 8. 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 reported in 1940 are for 1939.) Earnings and schooling regressions are restricted to ages 25–60. In the original, the notes to the table state that the literacy regressions are restricted to ages 16–60 while the text (p. 108) gives 15–45. I use the latter. As in the original, all regressions are run without a mean-reversion control (odd-numbered columns) and with (even-number columns.); this control is the product of age (equivalently, birth year) and a state-level measure of farm-worker wages in 1899 (Lebergott 1964).

The new results do not give great cause to doubt the original. 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, especially when including the mean-reversion control.

As discussed in section 2.3, I also run the Table V regressions while including Bleakley's full control set. This removes an asymmetry between the original Table V and Figure III, since the first does not cover full-controls specifications and the second does. Table 9 reports these regressions. For lack of better comparators, it juxtaposes the new with-full-control results with the original's without-full-control results. The most dramatic change wrought by the introduction of controls is a reversal of Bleakley's pessimistic findings for schooling (in the middle columns of the table). All of the negative impact coefficients for schooling become positive, and many of the positive ones grow. On the other hand, the positive impacts found on earnings of blacks largely disappear (last row).

TABLE 8. 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			
Controls 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	0.0286***	0.0154***	0.0234*	0.0197**	–0.0243	–0.0119	0.0037	0.0326	0.0158***	0.0065***	0.0115***	0.0028
× Years of exposure	(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	0.0254***	0.0161***	0.0219***	0.0189***								
× Years of exposure	(0.0044)	(0.0029)	(0.0063)	(0.0049)								
Hookworm infection rate	0.0023**	0.0024***	0.0022**	0.0024***								
× Years of exposure	(0.0009)	(0.0007)	(0.0009)	(0.0008)								
× Years of schooling												
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 9. REPLICATION OF BLEAKLEY (2007) TABLE V, ADDING FULL CONTROLS: 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			
Controls 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	0.0286***	0.0337***	0.0234**	0.0338***	–0.0243	0.1145**	0.0037	0.1144**	0.0158***	0.0108***	0.0115***	0.0117***
× Years of exposure	(0.0066)	(0.0094)	(0.0093)	(0.0093)	(0.0328)	(0.0568)	(0.0357)	(0.0558)	(0.0019)	(0.0021)	(0.0020)	(0.0025)
Observations		256806		256806		536029		536029		406200		406200
Panel B: Changing returns to schooling												
Independent variables												
Hookworm infection rate	0.0254***	0.0289***	0.0219***	0.0286***								
× Years of exposure	(0.0044)	(0.0078)	(0.0063)	(0.0076)								
Hookworm infection rate	0.0023**	0.0024***	0.0022**	0.0024***								
× Years of exposure	(0.0009)	(0.0007)	(0.0009)	(0.0007)								
× Years of schooling												
Observations		256806		256806								
Panel C: Estimates of hookworm × exposure for demographic subgroups												
Subsamples												
Males	0.0265***	0.0312***	0.0253***	0.0310***	–0.0690**	0.1194**	–0.0376	0.1225**	0.0108***	0.0016	0.0083***	0.0026
	(0.0056)	(0.0063)	(0.0080)	(0.0060)	(0.0326)	(0.0493)	(0.0347)	(0.0503)	(0.0018)	(0.0019)	(0.0019)	(0.0024)
Observations		189491		189491		266275		266275		201344		201344
Females	0.0322***	0.0433*	0.0157	0.0445*	0.0200	0.1093	0.0444	0.1059	0.0209***	0.0194***	0.0148***	0.0204***
	(0.0115)	(0.0227)	(0.0165)	(0.0235)	(0.0338)	(0.0685)	(0.0385)	(0.0651)	(0.0027)	(0.0040)	(0.0030)	(0.0042)
Observations		67315		67315		269754		269754		204856		204856
Whites	0.0293***	0.0333***	0.0232**	0.0337***	–0.0110	0.1290**	0.0164	0.1309**	0.0131***	0.0054**	0.0086***	0.0063**
	(0.0071)	(0.0111)	(0.0103)	(0.0112)	(0.0345)	(0.0615)	(0.0378)	(0.0624)	(0.0022)	(0.0023)	(0.0020)	(0.0028)
Observations		227359		227359		479501		479501		357414		357414
Blacks	0.0220***	–0.0067	0.0253**	–0.0068	0.1013***	0.1669*	0.0133	0.1733**	0.0314***	0.0170*	0.0262***	0.0181*
	(0.0072)	(0.0308)	(0.0103)	(0.0303)	(0.0387)	(0.0985)	(0.0461)	(0.0828)	(0.0065)	(0.0097)	(0.0063)	(0.0090)
Observations		29447		29447		56528		56528		48786		48786

All results derived as in Table 8 except that "new" regressions include the Bleakley (2007) full control set. Standard errors in parentheses, clustered by state of birth. $^{*}p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

4.2 Replicating and reanalyzing Figure III and Table VI: Multi-census retrospective cohorts

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

The RC regressions assess impacts on two IPUMS-provided measures of occupational standing. These are taken to proxy for income, a concept the U.S. Census did not begin directly tracking until the mid-20th century. Both proxies are constructed from variables that have figured in the census much longer, and so are available for all the census rounds used here. The occupational income score (OIS), introduced above in section, is an income index based on reported occupation. Duncan's (1961) socioeconomic index (SEI) blends in information about education level as well.

Like Figure II for the SCS regressions, 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 native whites aged 25–60.
2. Within each birth-year cohort between 1825 and 1965, fixed-effect dummies for each census year—or, equivalently, age—are partialled out of the occupational standing indicators.[19]
3. The two occupational standing indicators are then averaged by birth year, birth state, and census year, producing a three-dimensional panel.
4. Within each birth cohort, the two outcomes are regressed on H and optional state-level controls.
5. The resulting 141 coefficient estimates for H, $\hat{\beta}_t$, are plotted in Figure III.[20] Then, in Table VI, they are subject to time series analysis to assess whether Exp—the step function with the 19-year rise—is a strong predictor. These regressions are weighted by the square root 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 the spirit of the Frisch-Waugh-Lovell theorem. In principle, failure to partial the fixed effects out of the other right-side variables can cause their explanatory power to load onto those variables in an OLS regressions. In practice, this matters little because the other variables are cross-sectional, and so are nearly orthogonal to the census year effects. 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 cohort.
- Aggregating the data before the main analysis prevents controlling for micro-level demographic traits, which all the other Bleakley specifications do. In practice here, "traits" refers to sex; race is moot since the sample is restricted to whites and age effects are effectively removed by the partialling-out of birth year–census year effects in step 1.
- While weighting by the square root of cell size is evidently meant to improve efficiency by reducing heteroskedasticity, theory favors weighting simply by cell size. If the cohort-specific regressions in step 4

[19] Put otherwise, dummies for each census year–birth year combination are partialled out in the full sample, which is how Bleakley (2007) describes the process.

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

are individually homoskedastic, then the variance of the error term in each is inversely proportional to cell size.[21] Assuming that in the final-stage time series regression, this differential variance in the dependent variable, $\hat{\beta}_t$, carries over to the errors, then here error variance too is inversely proportional to cell size. This heteroskedasticity is reversed by weighting by inverse variance, i.e., by cell size. In symbols, if $\mathbf{Y}$ is a column vector holding the $\hat{\beta}_t$, $\mathbf{X}$ holds the right-side variables, and $\mathbf{W}$ is a diagonal matrix whose entries are cell sizes, then efficient weighted OLS for the time series is given by $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$.

- 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 AR() specifications than to help justify dropping them in my reanalysis.
- 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 treats the $\hat{\beta}_t$ as observed with perfect precision.

Of these concerns, only the last was pre-registered. (See section 2.1.)

After reconstructing the original figure and time series regressions, I implement an alternative approach that addresses or sidesteps all of the above critiques.[22] The alternative is merely to copy the practice in the rest of Bleakley (2007), directly fitting to the microdata. To compute the individual $\hat{\beta}_t$, I fit equation (2), above, to each birth cohort's microdata. The specification imports all demographic controls from the single-census RC regression that pertain, meaning fixed effects for each age-sex combination.[23] In fact, I consolidate all these regressions into a single, full-sample regression in which the δ_t are interacted with all other right-side variables. This facilitates the clustering of the standard errors by birth state, across cohorts, a step taken to combat serial correlation.

Then, to formally test whether Exp helps predict the $\hat{\beta}_t$, I estimate two versions of (1), which correspond to Models 1 and 2 in section 2.4. These regressions too are run on the microdata, not, as in Bleakley (2007), on the $\hat{\beta}_t$ derived in the previous step. The first version replaces the H_j in (1) with the three linear spline terms that generalize the functional form of Exp. Since Bleakley gives Exp a 19-year ramp-up, I give the "before" and "after" segments—the one's we imagine to be flat—19 years of coverage as well. To be precise, in the linear spline model, the regression replaces $H_j \times Exp_j$ in (1) with three terms:

$$H_j \times t, H_j \times \min(0, t-1891), H_j \times \min(0, t-1910) \tag{4}$$

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.

[21] More precisely, cell size times the variance estimate converges with sample size to the true variance, with probability 1.
[22] 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 AR() processes in the final stage and that for model without AR() terms, the omnibus OLS approach was appropriate, provided it could be made computationally practical.
[23] Below, when the data set is expanded to blacks, single-census RC controls involving race are also retained.

The second version of (1) retains $H_j \times Exp_t$ as a unitary term and instead echoes Bleakley (2007), Table VI, in controlling for polynomial trends in time. Each polynomial model of order d is fit to the full sample. The terms of interest, inserted in $\mathbf{z}$, are:

$$\{H_j \times t^r\}_{r=0,\ldots,d} \tag{5}$$

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

Under either model, this more direct approach to inference retains nearly all the substantive elements of the original: it combines data from many censuses while controlling within birth cohorts for census year effects. Improving on the original, it controls for sex, to address evolving gender roles in the labor market. It incorporates uncertainty in all steps into the final estimates. And it avoids the need to choose weights, square root or otherwise, for the $\hat{\beta}_k$. There is one econometric loss: putting all the estimation steps into an omnibus OLS regression makes it impossible to model the intermediate estimates, $\hat{\beta}_k$ as autoregressive. But as I argue above, this loss is not great. And it is offset by the more aggressive exploration of the step and polynomial models to test robustness, and by the clustering of the variance estimate by state, which should adjust for autoregressive serial correlation.

To start the application, Figure 4 attempts to imitate the original Figure III in data and method. It only departs substantively in adding (95%) confidence intervals for point estimates, which Bleakley (2007) Figure II also does, but Figure III does not. 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.

Table 10, below, does the same for Bleakley (2007) Table VI, reporting time series regressions on the dots in Figure 4. The first row of results is for the SEI regression without full controls, and corresponds to the upper left of Figure 4. The next row is for the bottom-left of Figure 4. And so on. Across the table, many of the new point estimates do not match the original ones that well. But the order of magnitude, sign, and significance are usually about the same.

Figure 5 updates Figure 4 by fitting to the expanded data set. This, recall from Table 1, adds 1860, 1930, and 2000 census data, and enlarges samples for other years. In addition, copying the rest of the paper, the sample is extended to blacks as well as whites. All the demographic controls in the single-census RC regressions involving race are now included. The time series are much less noisy now: this is obvious from a cursory comparison of Figure 4 and Figure 5, and becomes even clearer after one notes that the vertical ranges on the new graphs are narrower.

Figure 5 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? A gaze at Figure 5 suggests that the answer is "no." To formally test that interpretation, Figure 6 and Figure 7 fit the linear spline and polynomial models to the microdata. The figures retain the dots from the previous figure but, for legibility, drop the confidence intervals. For the same reason, for the polynomial model, fits of order 4 and 5 are not drawn.

The linear spline model fits in Figure 6 do not reject the null of no anomalous rise. Much as in Figure 3, in the bottom right of each graph are displayed *p* values for trend breaks in 1891 and 1910.

Figure 7 displays the polynomial fits. For the occupational income score (on the right of the figure), models of order 0, 1, and 2, in orange, green, and purple, mostly return positive and statistically significant coefficients on

Exp. However, moving to a cubic or higher-order fit greatly weakens this finding: see the red curves and the final *p* values in each right-side pane. And while we can expect that controlling for a time polynomial of high-enough order could deprive Exp of predictive value, even when Exp is part of the true model, the curves in Figure 7 do not appear to be overfitting the OIS data in this way. The ranges of highest curvature in the cubic model fit occur well outside the period of critical interest, 1891–1910. To the contrary, the lower-order models appear misspecified in the sense of Figure 1.

The polynomial fits for Duncan's socioeconomic indicator (left side of Figure 7) produce a less consistent pattern. Without controls, the impact estimates are statistically indistinguishable from zero. With controls, they tend to point to a positive impact of hookworm eradication, even controlling for a cubic function in time (p = 0.15). However, given the mild statistical significance of this finding, the weaker estimate under the cubic model for the occupational income score (lower right of Figure 7), and weak results in the corresponding linear spline fit (bottom left of Figure 6), this finding does not look very robust.

Table 11 provides more information on the polynomial model fits up to order 5, following the format of Bleakley (2007) Table VI. The coefficient on the treatment term $H \times Exp$ tends to stay positive and significant up to the quadratic model, lose significance in the cubic, and turn negative in the quintic. These results are roughly 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).

Perhaps a truer model would assume that hookworm exposure takes a long-term toll only when it occurs early in life. Without specifically mentioning helminths, Victora et al. (2008) suggest that health before age two may matter especially for later life. If so, then we should model Exp as rising much more suddenly, say, between the 1910 and 1912 birth cohorts. I have not formally tested that model, but the graphs to do not point to any such sudden change.

FIGURE 4. REPLICATION AND EXTENSION OF BLEAKLEY (2007) FIGURE III: RECONSTRUCTED DATA SET

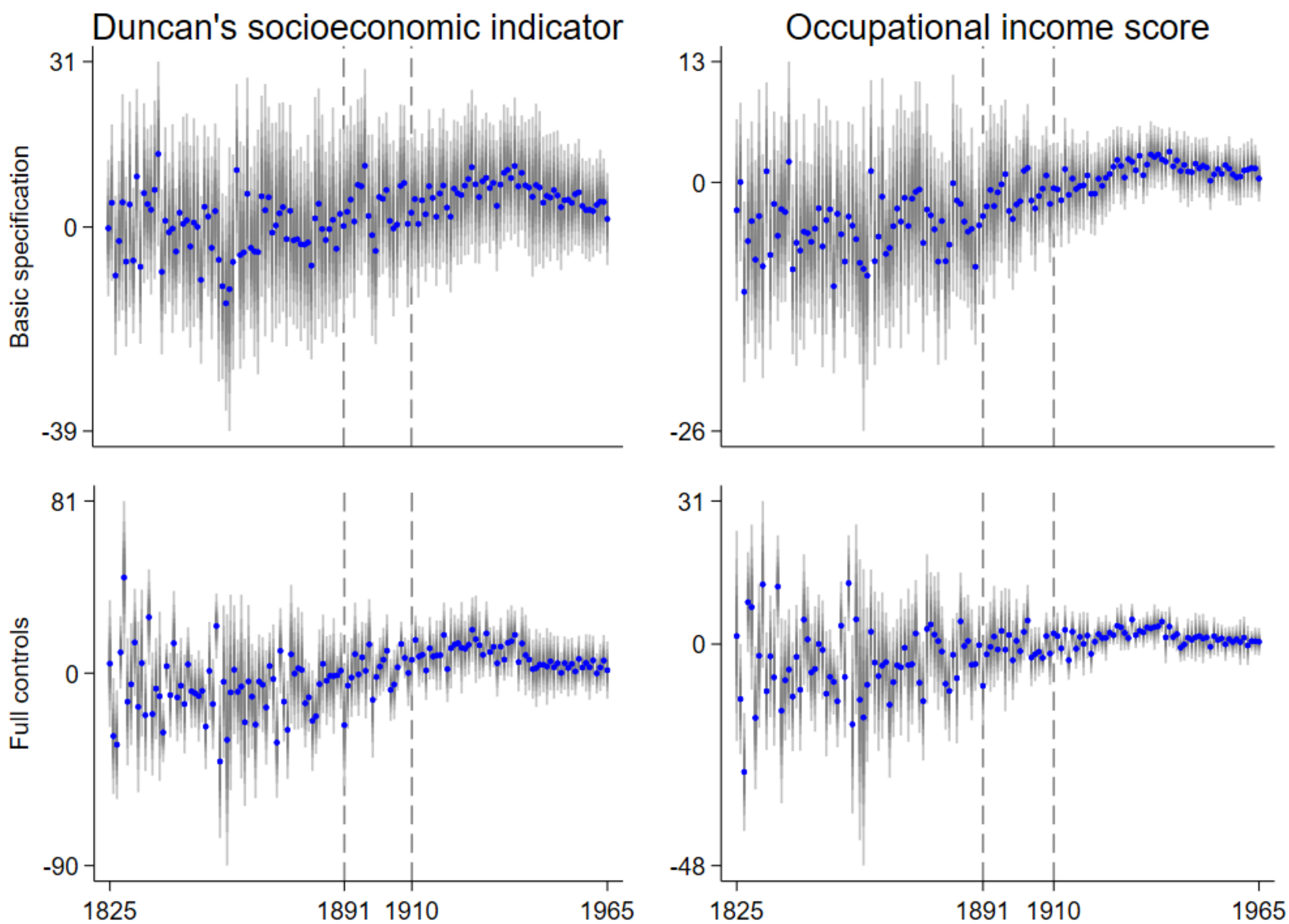


TABLE 10. REPLICATION OF BLEAKLEY (2007) TABLE VI: EXPOSURE TO RSC VERSUS ALTERNATIVE TIME-SERIES RELATIONSHIPS

		(1)		(2)		(3)		(4)		(5)	
Outcome	Controls	Original	New	Original	New	Original	New	Original	New	Original	New
Duncan's socioeconomic indicator	Basic	0.5352***	0.3407***	0.7566***	0.3363***	0.3928***	0.2965***	0.5983***	0.2927***	0.4858***	0.2473***
		(0.0418)	(0.0388)	(0.1069)	(0.0813)	(0.0520)	(0.0569)	(0.1124)	(0.0853)	(0.1282)	(0.0872)
Duncan's socioeconomic indicator	Full controls	0.5007***	0.7807***	0.8820***	1.0590***	0.3544***	0.8982***	0.6616***	1.2001***	0.7081***	1.2869***
		(0.0661)	(0.0989)	(0.1707)	(0.2310)	(0.0735)	(0.1334)	(0.1791)	(0.2643)	(0.1969)	(0.2852)
Occupational income score	Basic	0.3113***	0.3024***	0.2915***	0.1947***	0.2612***	0.2990***	0.2497***	0.1943***	0.1912***	0.1656***
		(0.0214)	(0.0206)	(0.0542)	(0.0483)	(0.0384)	(0.0431)	(0.0612)	(0.0492)	(0.0622)	(0.0471)
Occupational income score	Full controls	0.2623***	0.2849***	0.3732***	0.1907*	0.2346***	0.3340***	0.3393***	0.2078*	0.2742***	0.2731**
		(0.0339)	(0.0483)	(0.0858)	(0.1098)	(0.0438)	(0.0562)	(0.0960)	(0.1159)	(0.1007)	(0.1057)
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. $^{*}p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

FIGURE 5. REPLICATION AND EXTENSION OF BLEAKLEY (2007) FIGURE III: EXPANDED DATA SET

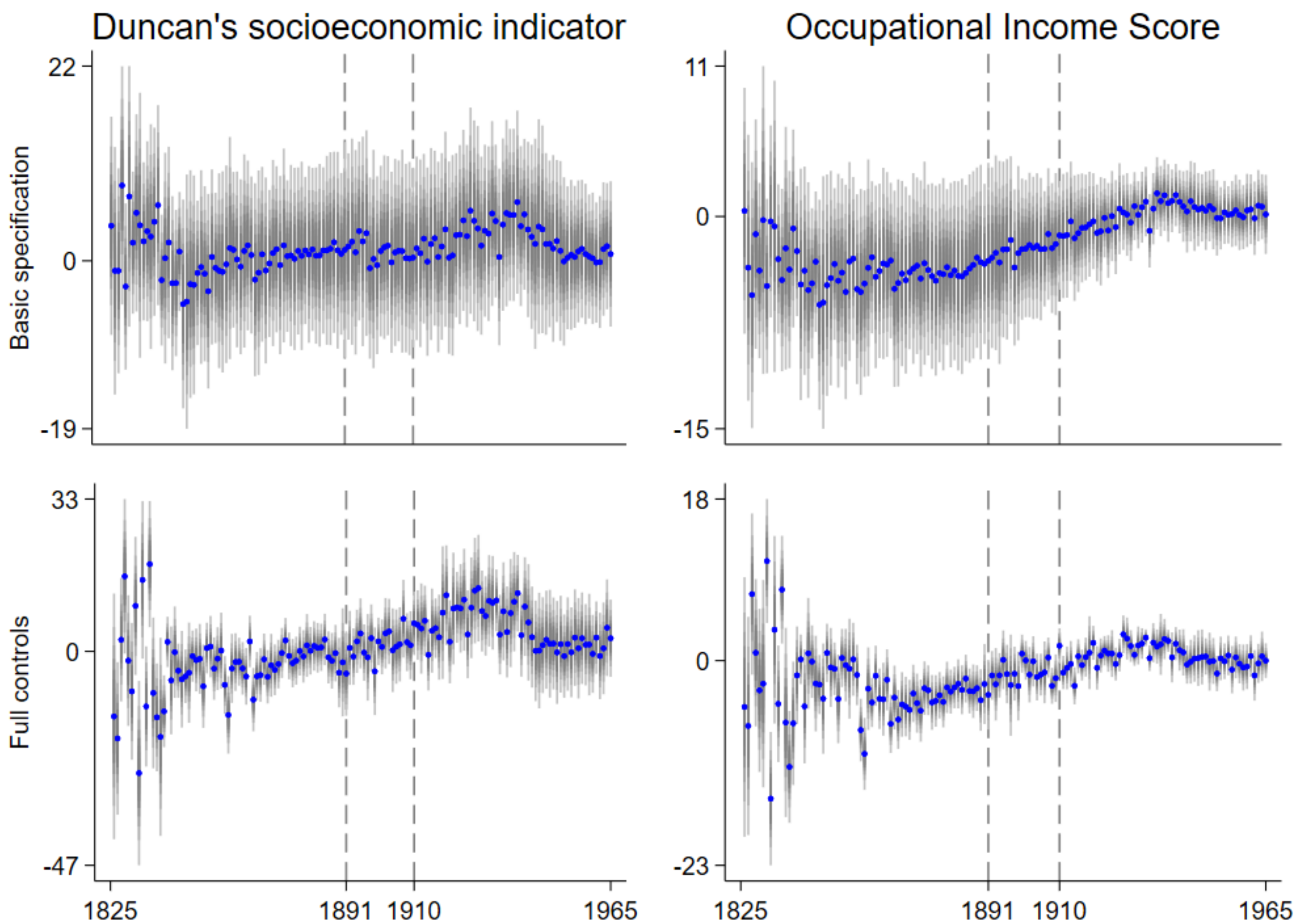


FIGURE 6. REPLICATION AND EXTENSION OF BLEAKLEY (2007) FIGURE III: MODEL WITH LINEAR SPLINE GENERALIZATION OF STEP FUNCTION

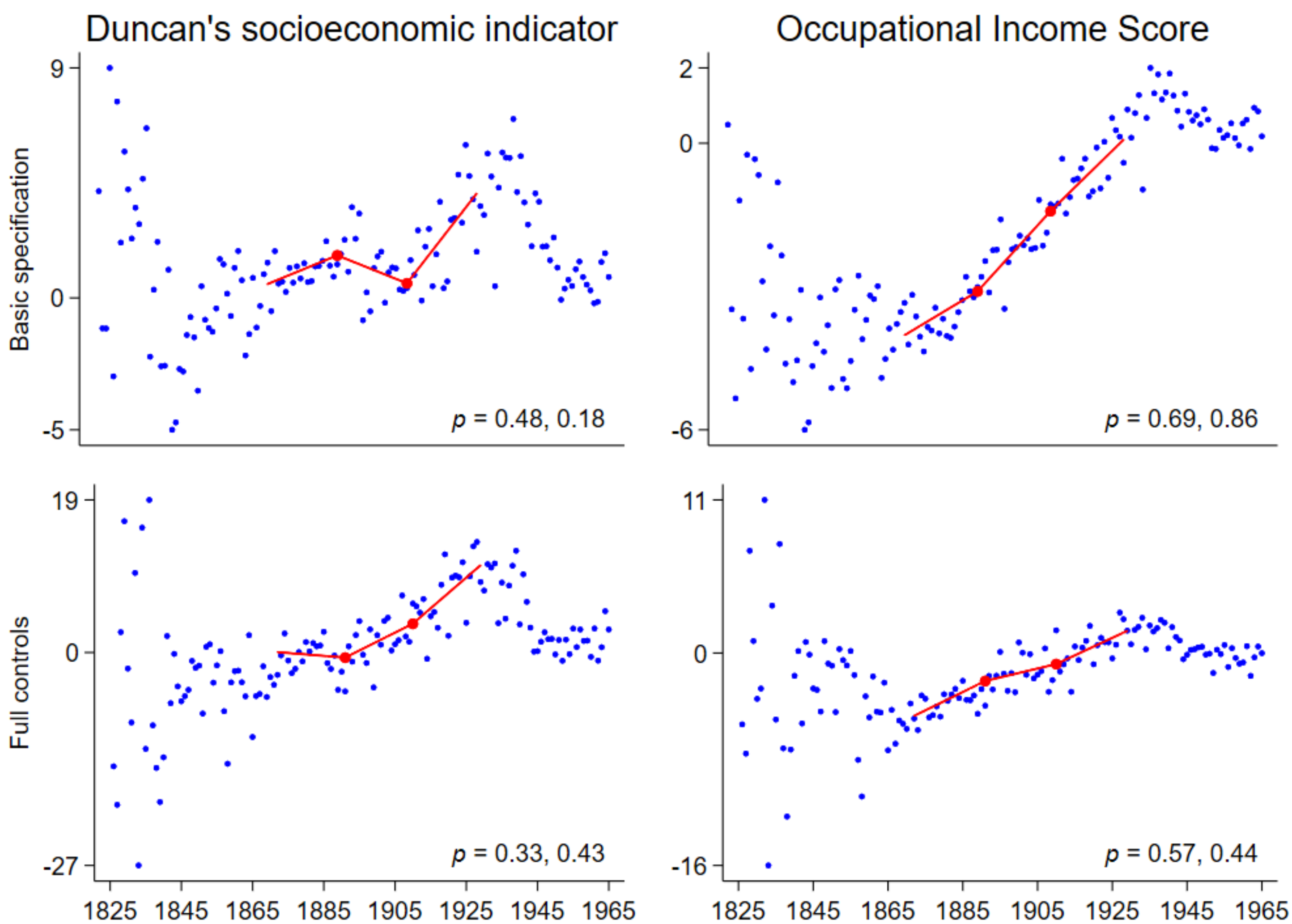


FIGURE 7. REPLICATION AND EXTENSION OF BLEAKLEY (2007) FIGURE III: MODEL WITH POLYNOMIAL TIME CONTROLS, FIT TO EXPANDED DATA SET

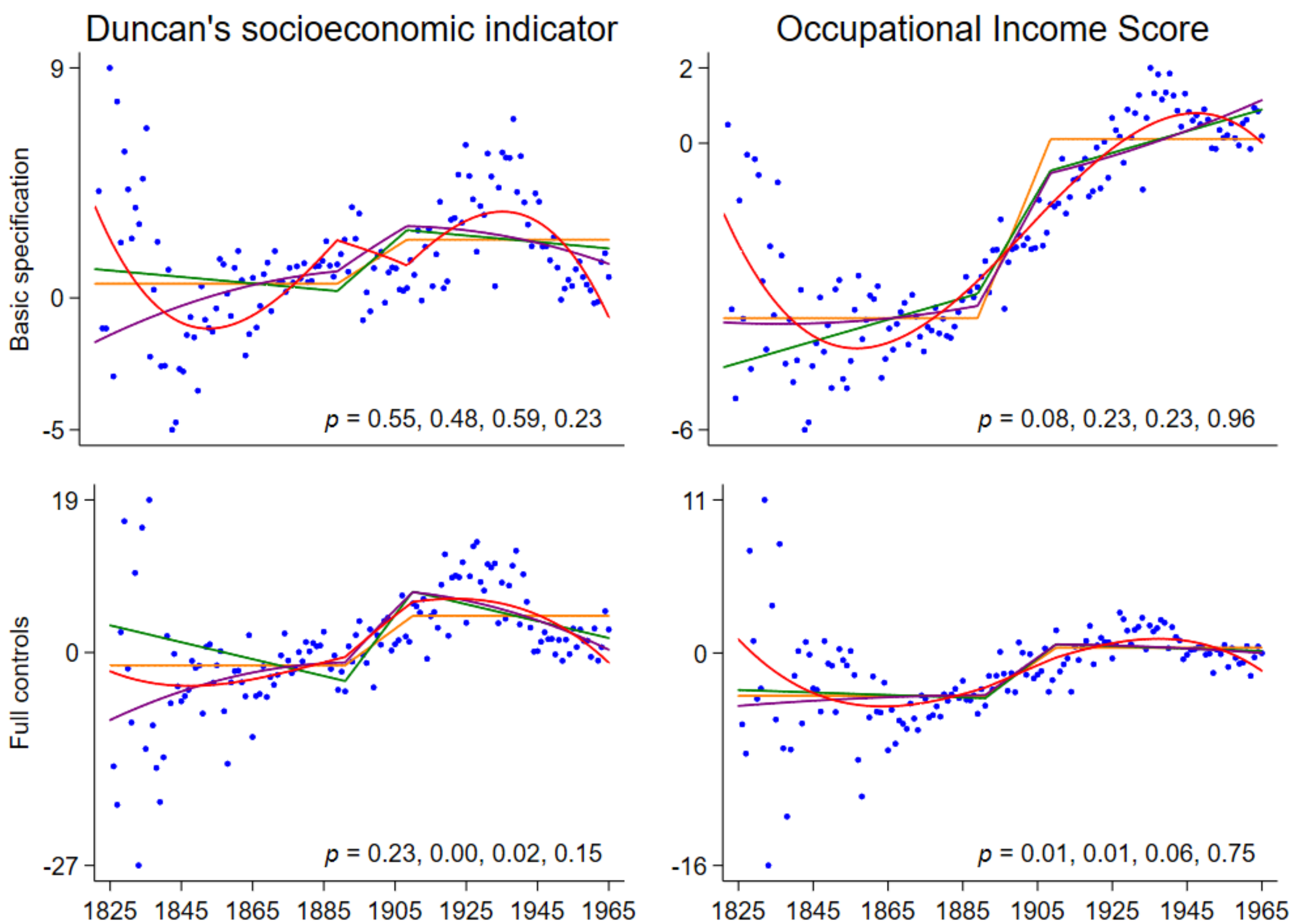


TABLE 11. REVISION OF BLEAKLEY (2007) TABLE VI: EXPOSURE TO RSC VERSUS ALTERNATIVE TIME-SERIES RELATIONSHIPS

Outcome	Controls	Coefficient on $H \times Exp$					
Duncan's socioeconomic indicator	Basic	0.0862 (0.1434)	0.1314 (0.1844)	0.0835 (0.1554)	–0.1958 (0.1617)	–0.1754 (0.1366)	–0.3905** (0.1460)
Duncan's socioeconomic indicator	Full controls	0.3268 (0.2681)	0.6975*** (0.2296)	0.4833** (0.1939)	0.2717 (0.1845)	0.1924 (0.1813)	–0.4565* (0.2369)
Occupational Income Score	Basic	0.2077* (0.1145)	0.1183 (0.0964)	0.1356 (0.1115)	0.0035 (0.0675)	0.0178 (0.0771)	–0.0499 (0.0636)
Occupational Income Score	Full controls	0.1882** (0.0704)	0.2207** (0.0828)	0.1993* (0.1038)	0.0284 (0.0902)	–0.0051 (0.0925)	–0.1894** (0.0889)
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. $^{*}p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

5 Conclusion

Bleakley (2007) identifies impacts from a variable that is the product of two factors: the geographic pattern of baseline hookworm burden and the sudden onset of the eradication campaign in the early 1910s. The first factor cannot be viewed as exogenous since it is a marker for climate and geography, and thus economic history. The second can be taken as credibly exogenous, but only in the short term. That is, the fact that eradication occurred between, say, 1850 and 1950, is part and parcel of the economic and scientific development of the United States. That it began in 1911 rather than 1906 or 1916 is more an accident of history. 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.

In my view, few of the regressions in Bleakley (2007) specify the model richly enough in the time domain to produce such evidence. Most effectively fit to a step function while controlling linearly for time. This model 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 these results do not appear robust. With reference to SCS regressions for short-term impact on school enrolment, the results do not appear consistently across measures of human capital investment and do not persist upon expansion of the census microdata set or inclusion of the full control set. Any rises in 1910–20 appear to have begun earlier. As for the RC specifications, expanding the census samples and applying formal tests for acceleration and deceleration at critical times leaves little evidence of positive, long-term impact on occupational standing.

That the original study concludes otherwise owes perhaps in part to its smaller census data sets, which do not allow the same precision in estimation. As well, certain results, such as the short-term association with schooling when controls are not included, seem disproportionately emphasized. And the original does not impose as heavy a burden of proof, evidently putting more weight on the tabulated regressions that I argue are easily misspecified.

Most of the revisions on which I base the judgment of fragility were not pre-registered. The exception is controlling for polynomials of order up to 5, which was implicitly pre-registered since Bleakley too ran such regressions. That said, all come from relatively obvious and non-arbitrary 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, GiveWell is publicly posting the full data and code for this replication. Until and unless the original data and code are 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.

References

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.

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

Bleakley, Hoyt. 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.

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

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*. http://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.

Duncan, Otis Dudley. 1961. “A Socioeconomic Index for All Occupations.” *Class: Critical Concepts* 1: 388–426.

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

Kofoid, Charles A., and John P. Tucker, "On the Relationship of Infection by Hookworm to the Incidence of Morbidity and Mortality in 22,842 Men of the United States Army," *American Journal of Hygiene*, 1(1921), 79-117.

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

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

Ozier, Owen. Forthcoming. “Exploiting Externalities to Estimate the Long-Term Effects of Childhood Deworming.” *American Economic Journal. Applied Economics*. 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 (July): CD000371.

Victora, Cesar G., Linda Adair, Caroline Fall, Pedro C. Hallal, Reynaldo Martorell, Linda Richter, Harshpal Singh Sachdev, and Maternal and Child Undernutrition Study Group. 2008. “Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital.” *The Lancet* 371 (9609): 340–57. doi:10.1016/S0140-6736(07)61692-4.

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.

Appendix: Additional figures

FIGURE A 1. EXTENSION OF BLEAKLEY (2007) FIGURE II TO ADULT OUTCOMES: EXPANDED DATA

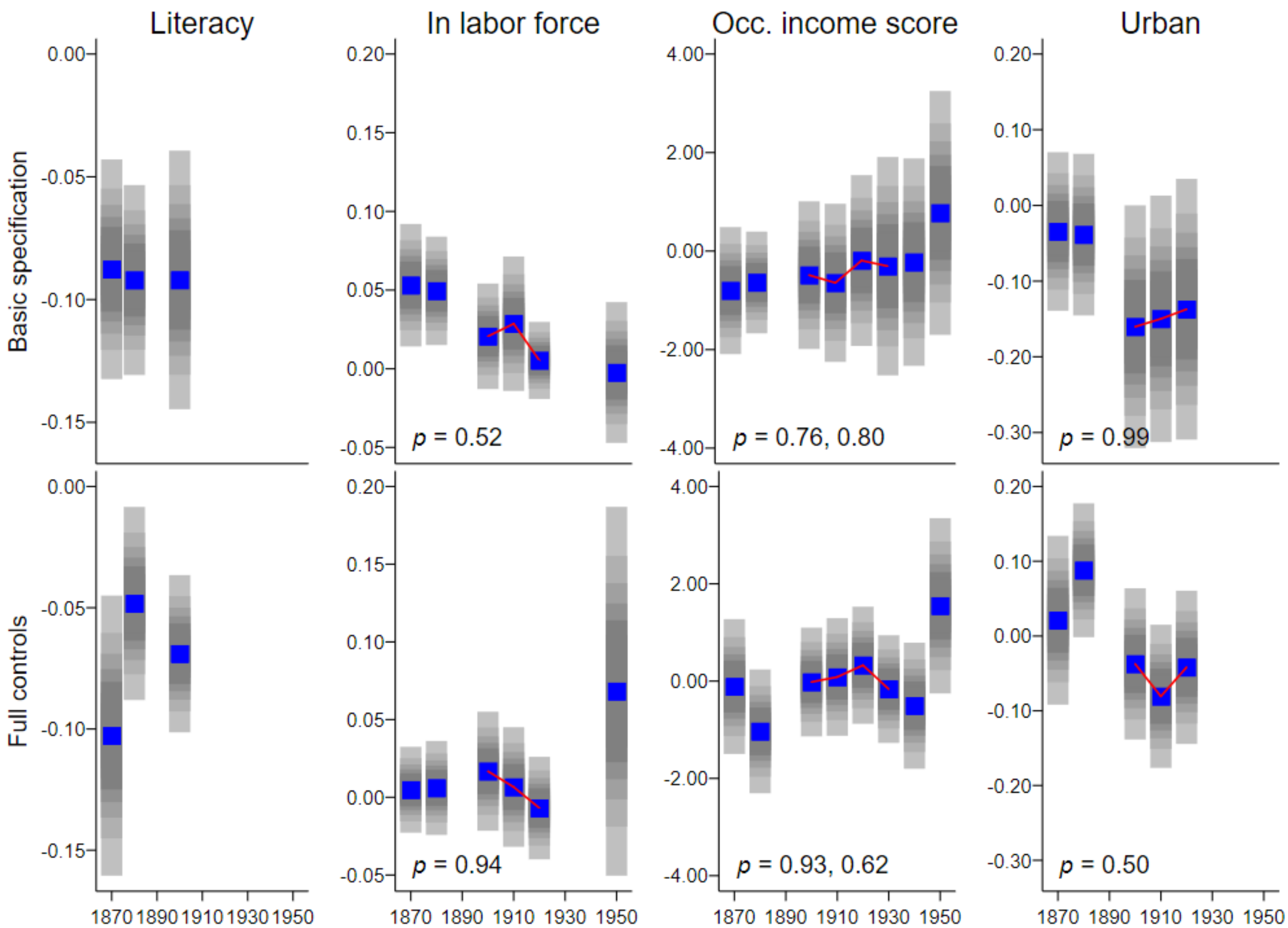


FIGURE A 2. TIME SERIES VARIANT OF BLEAKLEY (2007) FIGURE II, WITH SEPARATE REGRESSIONS FOR BELOW- AND ABOVE-40%-PREVALENCE SAMPLES, EXPANDED DATA SET