Malaria Eradication in the Americas. A replication study of Bleakley (*American Economic Journal. Applied Economics*, 2010)

David Roodman[*]

International Journal for Re-Views in Empirical Economics, Volume 2, 2018-4, DOI: 10.18718/81781.8

JEL: I18, O15
Keywords: malaria; 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 ipums.org. Some of these data sets—the 100% U.S. census samples for 1910 to 1940 and the non-U.S. data sets—are not licensed for public redistribution. The original data and code provided by Bleakley (2010) can be downloaded from aeaweb.org/aej/app/data/2008-0126_data.zip. The redistributable 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). Malaria Eradication in the Americas. A replication study of Bleakley (American Economic Journal. Applied Economics, 2010). *International Journal for Re-Views in Empirical Economics*, Vol 2(2018-4). DOI: 10.18718/81781.8

Abstract

Bleakley (2010) finds that large-scale campaigns in the 20th century to eradicate malaria were followed by income gains for those native to historically endemic areas. I perform a pre-registered reanalysis and find these results to be largely robust. Malaria eradication efforts indeed appear to have been followed by anomalous income gains for natives of historically malarial areas of Brazil, Colombia, Mexico, and perhaps the United States. This supportive finding diverges from that of a separate, parallel reanalysis of Bleakley (2007), a study that finds long-term benefits from a hookworm eradication campaign in the United States.

[*]Senior Advisor, GiveWell, 182 Howard St., #208, San Francisco, CA, 94105 (e-mail: david.roodman@givewell.org).
The author thanks an anonymous referee for constructive comments, Christian Smith for help with data collection, and Zachary Tausanovitch for help with data collection and database construction. GiveWell and its spin-off, the Open Philanthropy Project, have no financial interest in the topic of this paper that constitutes a conflict of interest.

Received December 26, 2017; Revised June 11, 2018; Accepted November 7, 2018; Published December 14, 2018.

1 Introduction

Bleakley (2007) and Bleakley (2010) are important contributions to the literature on the long-term economic impacts of public health interventions. Both find that large-scale campaigns in the 20th century to eradicate a parasitic disease—hookworm and malaria, respectively—were followed by relative income gains for those native to historically endemic areas. These findings suggest that campaigns to improve public health can have substantial long-term impacts on human capital investment and productivity. The first (hookworm) study is set in the United States, the second in the United States, Brazil, Colombia, and Mexico. Roodman (2018) replicates and reanalyzes the first, and ultimately questions its conclusion, arguing that no historical discontinuities clearly coincide with the hookworm eradication campaign. The present paper brings the same set of techniques to Bleakley (2010)'s study of anti-malaria campaigns.

As a *replication*, this paper returns to primary sources to reconstruct all the variables for the U.S. impact assessment. For Brazil, Colombia, and Mexico, it likewise reconstructs the outcome variables, but, for practical reasons, not the treatment proxies or controls. The paper uncovers some coding errors in the original, but these do not appear to greatly affect results. As a *reanalysis*, the paper introduces (pre-registered) innovations: improving the outcome measures by incorporating the denser samples of census microdata now available; and applying formal and graphically informed inference to time series patterns.

The reanalysis tends to corroborate Bleakley (2010). Adult earnings as a function of birth year rose with anomalous speed in historically malaria-burdened regions about when the first babies were born who would spend part of their childhoods in post-eradication regimes. And convergence decelerated as the last of these babies were born—that is, as the transition from pre- to post-eradication regime completed. The corroboration is less certain for the United States than for Brazil, Colombia, and Mexico. Perhaps childhood malaria exposure mattered less for lifetime earnings in the U.S. Or perhaps the asserted natural experiment was weaker there, because the country operated at the technological frontier of malaria eradication: public health innovations that disseminated over the decades in the U.S. before World War II were gathered into sudden big pushes in poor countries after.[1] The fingerprint of impact is also less clear for human capital accumulation, as measured by literacy in adulthood and years of schooling completed. That result somewhat contradicts Bleakley (2010), which perceives indications of impacts on literacy, if not schooling.

This replication effort speaks not only to the impact of public health interventions on economic development. It also offers lessons on how journals archive data and code. The data availability policy of *AEJ: Policy*, which published Bleakley (2010), requires authors to provide "the data, programs, and other details of the computations sufficient to permit replication."[2] Hoyt Bleakley

[1] In addition to the Rockefeller Foundation–led effort beginning in the early 1920s, which is the basis in Bleakley (2010) for perceiving a natural experiment in the U.S., there came other influences: the hookworm eradication campaign beginning in the early 1910s may have helped by lessening the immonusuppressive effects of this disease in the population; in the late 1930s, the Works Progress Administration devoted 2 percent of its budget in the South to malaria eradication; at the same time, the Tennessee Valley Authority built dams that created new bodies of standing water in Alabama, Mississippi, and Tennessee, which may have *slowed* progress against malaria; also during the Depression, the Agricultural Adjustment Administration made payments to take farmland out of production, which may have stimulated migration out of malaria-endemic areas; and DDT spraying was introduced in the early 1940s. See Barreca, Fishback, and Kantor (2012).

[2] web.archive.org/web/20171101092538/https://www.aeaweb.org/journals/policies/data-availability-policy

appears to have complied with this policy as it has normally been implemented, providing data and code to the journal's website.[3] Yet in two important respects the paper's results are impossible to exactly replicate. The figures, which are no less important than tables for inference, cannot be precisely reproduced, because the public code does not generate them. Lack of public code for figures appears to be the norm for the American Economic Association journals. Also, neither the primary data nor the code that transforms it into analysis data are included—as again appears to be the norm—so one cannot easily reconstruct the chain from primary sources to final conclusions.[4] In these ways, the AEA archives fall short of their purpose of making research transparent and replicable.

Section 2 of this paper describes the Bleakley (2010) research designs. Section 3 explores some cross-cutting themes in the replication and reanalysis. Section 4 reports on the (partial) reconstructions of the data sets. Section 5 replicates and reanalyzes the time series results. Section 6 concludes.

2 Designs

The Bleakley (2010) specifications combine up to three sorts of variables:

- Cross-sectional variables observed once per geographic unit—for example, per Brazilian state or Colombian *municipio*. These include indicators of pre-eradication malaria mortality or malaria ecology (M), as well as controls.
- Variables indexed by individual, time, place and built from census microdata, including measures of schooling, literacy, and income. All microdata come from the Integrated Public Use Microdata Series (IPUMS; Ruggles et al. 2015; Minnesota Population Center 2017).
- A pure time series indicator for potential exposure to a national eradication campaign (Exp). Only the panel regressions, described shortly, include Exp explicitly. In an approach akin to difference-in-differences, these regressions interact Exp with M to form the treatment proxy, while effectively controlling for Exp and M individually.

Of the two components of the $Exp \times M$ treatment proxy, the second is a marker for geography and therefore potentially for local economic history. While *external* to the causal pathways from malaria eradication to the outcomes of interest, it is not very credibly exogenous, in that historical factors can simultaneously influence treatment and outcome. The other component, Exp, is more plausibly exogenous in the short-term than the long-term. That is, it is not an accident of history that these campaigns occurred in the 20th century rather than the 19th or 21st. More accidental perhaps is that they took place precisely when they did, rather than a few years earlier or later. Thus, as in an interrupted time series design, the results that can most compellingly demonstrate causality will derive from changes over short timeframes.

All the Bleakley (2010) estimators begin by averaging an outcome Y within census year–birth year–birthplace cells, with the dimensions indexed by c, t, j, respectively; this gives a set of values

[3] aeaweb.org/aej/app/data/2008-0126_data.zip
[4] See also Glandon (2011).

$\bar{Y}_{ctj}$. These are next demeaned nationally within each census year–birth year group, yielding $\widetilde{Y}_{ctj}$; that is, census year–birth year fixed effects are removed. The $\widetilde{Y}_{ctj}$ are then fit in regressions. A disadvantage of this preprocessing is that the imprecision of the initial demeaning, which is a preliminary estimation step, is not factored into the standard errors from the main estimation step.

Bleakley (2010) first fits cross-sectional long-difference regressions, with the model

$$\Delta\widetilde{Y}_j = M_j\beta + \mathbf{x}_j'\boldsymbol{\gamma} + \epsilon_{ij} \tag{1}$$

β is the parameter of interest.[5] $\mathbf{x}$ is a set of controls, ϵ_{ij} is a mean-zero random error, and $\Delta\widetilde{Y}_j$ is the change in the average value of $\widetilde{Y}_{ctj}$ for area j from the "before" to the "after" period. The "before" period ends in 1890 in the United States, and in 1940 in Brazil, Colombia, and Mexico. These cut-offs are chosen to assure that all children born in the "before" period would have reached adulthood by the campaign, and so would have experienced no campaign-induced reduction in childhood malaria exposure. The "after" period starts when the eradication campaign is taken to have commenced—1920 in the United States, 1957 in the Latin countries. People born after these dates are considered to have grown up fully within the post-eradication regime. Individuals born in the gap between the two periods do not figure in these regressions.

The long-difference regressions, reported in Bleakley (2010) Tables 1–3, show that most measured outcomes improved faster in places with high pre-eradication malaria burden. These relative rises constitute circumstantial evidence that eradication efforts delivered substantial benefits. However, as Bleakley (2010, p. 13) points out, the regressions do not speak to the historical distinctiveness of the rises. Perhaps these trends began too early or continued too long for the malaria eradication campaigns to naturally explain them.[6]

The Bleakley (2010) panel regressions look more sharply at timing. To do so, they define the exposure variable Exp as the fraction of childhood spent in the post-eradication regime, as a function of birth year. As a pure time series variable, Exp takes the same value regardless of the historical malaria burden of one's birthplace. According to the Bleakley (2010) text, childhood is taken to last 21 years. This makes Exp a piecewise-linear "step" function with a 21-year rise. In the Latin countries, for example, Exp is 0 through 1936, then rises linearly until it reaches 1 in 1957, and then runs flat again.

The panel regressions fit

$$\widetilde{Y}_{ctj} = (Exp_t \times M_j)\beta + \mathbf{x}_{tj}'\boldsymbol{\gamma} + \delta_c + \delta_t + \delta_j + \epsilon_{ctj} \tag{2}$$

β remains the parameter of interest. The δ_c, δ_t, and δ_j are dummy sets for census year, birth year, and birthplace, with the δ_t and δ_j obviating the need to include Exp_t and M_j as controls. The controls $\mathbf{x}_{tj}$ are not true panel variables in the sense of being observed in primary sources in multiple times in multiple places. Rather, all are products of pure cross-sectional and pure time series

[5] These can also be viewed as two-period panel regressions in which Exp is a dummy for the second period, $M \times Exp$ is the treatment proxy, and M and Exp are effectively controlled for through dummy sets for place and year of birth.

[6] Bleakley (2010, p. 13) suggests that because they apply to data aggregated over time, the long-difference regressions have the advantage of avoiding high-frequency serial correlation. However, the Bleakley (2010) panel regressions also address serial correlation, by clustering standard errors by place of birth.

variables. For example, the Bleakley (2010) "full controls" panel regressions include interactions between Exp and geographic controls relating to health and education.

Regressions based on (2) can be viewed as testing whether the step function Exp is a strong explanator for the temporal evolution of the spatial association between baseline malaria burden M and the outcome *Y*. The model will fit well if the association takes a low (potentially negative) value among cohorts born well before the campaign, begins to rise steadily among those born late enough to still be children during the campaign, and then plateaus again among people born after the campaign.

However, fitting the model can still generate a false positive if such convergence begins well before or extends well after the dates implied by the construction of Exp—and is in fact caused by other forces. Regressions in such cases could estimate β as being statistically different from zero, and create the spurious impression that Exp is a good explanator for what are longer-term trends.

Bleakley (2010) takes several steps to rule out such possibilities. All the Bleakley (2010) regressions include measures of initial conditions in order to control for mean reversion. Some introduce state- or *municipio*-specific time trends, linear or quadratic. These measures suffice *if* the augmented models largely capture the ambient time trends. But in general, we do not know the functional form for major ambient trends. And it is hard to judge how close the models come only by viewing tabulated estimates of β.

Bleakley (2010)'s graphical time series approach can give more insight into ambient trends. It runs a version of (2) within each (t-indexed) birth cohort:

$$\widetilde{Y}_{ctj} = M_j\beta_t + \mathbf{x}'_{tj}\boldsymbol{\gamma}_t + \delta_{ct} + \epsilon_{ctj} \tag{3}$$

By dropping the single treatment term $Exp \times M$ in favor of M_j and giving it cohort-specific coefficients, this equation removes any modeling restriction on the temporal evolution of the cross-section association between M and the outcome, an association represented by β_t. The β_t can be graphed for visual inspection of long-term trends. And they can in turn be subjected to formal inference. The cost of the modeling change is the loss of the fixed geographic effects δ_j, which cannot be identified in these cross-geography regressions.

In studying hookworm eradication, Bleakley (2007, Table VI) uses time series regressions to perform inference on whether Exp is a determinant of the β_t. In contrast, Bleakley (2010) discusses the evolution of the β_t only informally. I resurrect and revise the Bleakley (2007) approach and apply it to malaria eradication, just as Roodman (2018) does for the Bleakley (2007) hookworm study.

This revised time series approach begins by fitting (3) directly to census microdata, as in most of the Bleakley (2007) hookworm study, rather than to nationally demeaned, cell-aggregated outcomes, $\widetilde{Y}_{ctj}$. This change brings three benefits. First, moving to microdata sidesteps the arguable choice in Bleakley (2010) to weight observations by the square root of cell size instead of cell size.[7]

[7]Weighting by the square root of cell size is evidently meant to improve efficiency by reducing heteroskedasticity. But theory favors weighting simply by cell size. The variances of the cell-averaged values $\widetilde{Y}_{ctj}$ are inversely proportional to cell size. Assuming that this inverse law carries over to the $\widetilde{Y}_{ctj}$ and ϵ_{ctj}, the heteroskedasticity is reversed by weighting

Instead, one weights individuals by the IPUMS-provided sampling weights. Second, the move allows one to incorporate individual-level demographic controls. As the regressions are carried out here, this amounts to including a dummy for sex in the expanded-sample regressions, since they add women (see section 3.2); and likewise for race in the expanded U.S. regressions, which also add blacks. Bleakley (2007) uses both dummies too. (Within birth cohorts, controlling for fixed census round effects effectively controls for age already.) Since labor market participation and labor market outcomes evolved in distinctive ways for men and women over the study period, the cohort-specific gender dummies should removing these trends and thereby improve precision in the search for the fingerprint of malaria eradication. In the U.S. context, the same goes for the added race dummies. The third benefit of fitting to microdata is that it allows one to effectively merge the Bleakley (2010) preprocessing step—national demeaning—into the main estimation step, to assure that standard errors reflect imprecision in both steps.

Formally, I rewrite the cohort-specific cross-section model (3) as

$$Y_{ictj} = M_j\beta_t + \mathbf{z}'_{ictj}\boldsymbol{\alpha}_t + \mathbf{x}'_{tj}\boldsymbol{\gamma}_t + \delta_{ct} + \epsilon_{ictj} \tag{4}$$

The new index i identifies individual census observations. The δ_{ct}, dummies for each census year–birth year combination, effect the Bleakley (2010) preprocessing. The new variable set $\mathbf{z}$ holds individual-level traits observed in censuses; they may take different coefficients for each birth cohort since $\boldsymbol{\alpha}$ is indexed by t.

The regressions (4) are implemented for all birth cohorts at once via a single, full-sample regression in which time dummies δ_t are interacted with all the independent variables. This facilitates clustering the standard errors by birthplace, across birth cohorts, to mitigate serial correlation. The resulting estimates of the β_t are plotted.

To formally test whether Exp helps predict the β_t, I then estimate three versions of (4), all of which impose restrictions on the structure of the β_t, much as in the Bleakley (2010) panel specification. Instead of allowing the β_t to take independent values, I again estimate a single coefficient on $Exp_t \times M_j$:

$$Y_{ictj} = (Exp_t \times M_j)\beta + \mathbf{z}'_{ictj}\boldsymbol{\alpha}_t + \mathbf{x}'_{tj}\boldsymbol{\gamma}_t + \delta_{ct} + \epsilon_{ictj} \tag{5}$$

The first test of the explanatory contribution of Exp echoes Bleakley (2007, Table VI) in introducing controls for polynomial trends in time. The terms of interest, inserted in $\mathbf{x}$ in (5), are:

$$\{M_j \times t^r\}_{r=0,\ldots,d} \tag{6}$$

d ranges up to 5 because Bleakley (2007, note 25), reports testing up to quintic 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

$$BIC = k\ln N + N(1 + \ln\tau + \ln MSE) \tag{7}$$

by inverse variance, i.e., cell size. In symbols, if $\mathbf{Y}$ is a column vector holding the $\widetilde{Y}_{ctj}$, $\mathbf{X}$ holds the independent variables, and $\mathbf{W}$ is a diagonal matrix whose entries are cell sizes, then Aitken's efficient generalized least squares estimator is $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$. The Bleakley (2010) code performs $(\mathbf{X}'\mathbf{W}^{1/2}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}\mathbf{Y}$.

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. As explained in Roodman (2018), I set N to the number of birth years in the sample rather than the number of observations in the individual-level microdata, because of clustering. E.g., for the U.S. we have 141 birth cohorts from 1825 to 1965, so N=141.[8]

The second version of (5) used to formally test the explanatory power of Exp does not include the terms in (6). Instead, it introduces three linear spline terms to generalize the step-like functional form of Exp. This "fixed-spline" model loosens the restriction that Exp is flat before and after the transitional ramp-up period, and allows a formal test of slope change at the corners—that is, of whether relative progress in high-malaria regions accelerated and decelerated when predicted. Since the Bleakley (2010) text ascribes a 21-year ramp-up phase to Exp, I give each spline segment 21 years of coverage. To be precise, the spline model regression replaces $Exp_t \times M_j$ in (5) with the three terms:

$$t \times M_j, \min(0, t - Campaign\ year - 21) \times M_j, \min(0, t - Campaign\ year) \times M_j \tag{8}$$

where $Campaign\ year$ is 1920 for the United States and 1957 for Brazil, Colombia, and Mexico; and $\min(\cdot)$ is the minimum function. The sample is restricted to those born between 21 years before the first kink and 21 years after the second, for a range of up to 63 years, data availability permitting.

Giving each segment a length of 21 years reflects an arbitrary choice, but one intended to be minimally so. In general, lengthening the outer segments would give more weight to long-term developments, in a context where the plausibly exogenous variation is short-term. For example, in the United States, if the β_t fell steadily between 1830 and 1865 and then symmetrically recovered between 1865 and 1900, extending the first spline segment from 1899 back to 1830 might give it a fairly flat slope in the best fit, obscuring the steady rise that begins well before the first hypothesized kink point. On the other hand, shortening the outer segments reduces statistical power. Giving the outer segments the same 21-years span as the inner one therefore seems like an appropriate compromise.

A disadvantage of the fixed spline model is implied by its name: it chooses the kink dates *a priori*. Yet the dates come from an impact model that, while reasonable, could be inaccurate, as Bleakley (2010) points out. The benefits of malaria eradication may not accrue in exact proportion to the fraction of one's first 21 years spent in the post-eradication regime. For this reason, the last modeling approach estimates kink dates from the data along with all of the OLS coefficients. It uses a mean squared error criterion much as in the Bai and Perron (1998) approach to identifying structural breaks.[9] 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 data. The method does not easily support formal inference with respect to the kink dates since they are discrete parameters. And the fit may be drawn to large structural breaks whose timing could not be explained by malaria eradication campaigns. Still, the results are informative as to whether trend

[8] 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.

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

shifts temporally associated with anti-malaria campaigns are major features of the historical record.

As tools for testing the explanatory value of Exp, the three models have advantages and disadvantages. The polynomial models carry some risk of generating spurious results: the low-order polynomials may underfit, endowing false explanatory power on Exp; and the high-order polynomials may overfit.[10] The fixed-spline model provides a focused and intuitive test of whether relative gains in income and human capital broke from ambient trends with timing naturally explained by malaria eradication efforts. Yet the model is somewhat arbitrarily moored to specific kink dates. The flexible-spline model solves that problem, but introduces a new risk, that the model will be drawn to large trend breaks unrelated to malaria programs.

The upshot of these conceptual difficulties is that one should not take any one of the regression results as definitive, and instead exercise judgment in blending all.

3 Themes in the replication and reanalysis

3.1 Pre-analysis plan

I registered a pre-analysis plan for this paper with the Center for Open Science.[11] I did not allow the plan to limit the analysis. But I found little cause to deviate from the plan because I had nearly completed the replication and reanalysis of the closely related Bleakley (2007), and this strongly informed the plan for Bleakley (2010).

The plan sets out several steps, which are listed here with commentary:

- *"Searching the figures and tables for asymmetries, such as one set of regressions being conducted at the individual level and another at the geographic level, and, where appropriate and practical, testing robustness of the results to copying specification choices from one to the other."* Two arguable asymmetries are exploited. The U.S. regressions are for whites only while the Brazil, Colombia, and Mexico ones include all races, even as most of the Bleakley (2007) U.S. regressions also include blacks. Here, blacks are added. (More on this issue just below.) Also, the Bleakley (2010) long-difference regressions apply to more outcomes than do the panel regressions. The long-difference regressions include, for example, literacy and years of schooling in the Latin countries. The reanalysis treats all the outcomes symmetrically.[12]
- *"Formally testing whether the curve fits in figure 4 are statistically significant, and whether those results are robust to inclusion [of] controls for linear or higher-order trends in time (up to order 5)."* The "curve fits" are the graphical superpositions of Exp on the estimates of β_t in Bleakley (2010) Figure 4. The formal methods for assessing this fit are discussed just above. All are copied from the replication and reanalysis of Bleakley (2007).
- *"Testing robustness of the above to 1) a switch from data aggregated by census year, birth year, and birth state to individual-level data; 2) expansion to blacks and women; and 3) incorporation of controls for race, sex, census year, and all their interactions."* The move to microdata

[10] Bleakley (2010, p. 24) warns that "horse-racing the exposure with second-degree trends across cohorts is a more difficult test to pass" in the data sets from Latin America, with their shorter time spans.

[11] See osf.io/h98yf.

[12] The Bleakley (2006) working paper does also include panel regressions and graphs for these additional outcomes.

is motivated just above. All these choices mimic the majority of the Bleakley (2007) specifications. Bleakley (2010) argues that restriction to men makes for a cleaner analysis since "their labor-force participation is higher and more consistent across the wide swath of years" (p. 11). Bleakley (2010, note 7) makes a similar argument for excluding blacks, but here the paper is not quite as internally consistent. The Latin American samples include all races, if only because "race was not measured consistently in the Latin America sample" (Bleakley 2010, note 7). However, the present reanalysis is premised on the view that the most plausibly exogenous identifying variation comes the specific timing of eradication, which argues for maximizing power to detect developments over shorter timespans, at the expense of longer-term comparability. Even if distinctive over the long run, trends for blacks and women could be expected to kink in the same ways as for white men. Or if they do not, this information informs a fuller assessment of the impacts. I follow Bleakley (2007) in adding controls for the dimensions of demographic expansion.

- *"When working with aggregate data, testing robustness to weighting by cell size rather than the square-root thereof."* Weighting by cell size—the number of primary observations behind each aggregated observation in the analysis data set—should better assure efficiency in the face of heteroskedasticity.[13] However, this point is largely moot since I work mainly with microdata.
- *"Testing robustness to the incorporation of newer and larger census samples from IPUMS."* This is done, as discussed in the next subsection.
- *"In the case of the U.S., testing robustness to switching as much as possible to the data set recently reconstructed from primary sources [for Roodman (2018)] in order to replicate Bleakley (2007)."* This is done.

3.2 Expanded census samples

The IPUMS census microdata collection has expanded steadily over the years: in countries and census rounds included and, at least for the United States, in the size or "density" of samples digitized. Bleakley (2010) largely does not specify the densities of the samples it uses. But they can be estimated from the reported download dates and the history of certain ipums.org pages at archive.org.[14] Table 1, column 1, shows my estimates.

I test robustness by switching to newer, larger IPUMS samples. For the United States, the expansion introduces data for 1860, 1870, and 1930. It raises the density from 1 percent to 5 percent in 1900 and 1960, and to 100 percent for 1910–40. Column 2 of Table 1 provides more detail. As just noted, in expanding the samples, I add women and, in the U.S. case, blacks. The Latin American IPUMS samples have not become denser since Bleakley accessed them. But more have become available, and are incorporated here: Brazil 2010; Colombia 2005; and Mexico 1995, 2010, and 2015.[15]

[13] See note 7.

[14] Bleakley (2010) reports last obtaining U.S. data from IPUMS on November 14, 2005, and last accessing Brazil, Colombia, and Mexico data in April 2006. See the change log at usa.ipums.org/usa-action/revisions and the archive.org history of ipums.org/usa/sampdesc.html, international.ipums.org/international/sample_designs/sample_designs_br.html, international.ipums.org/international/sample_designs/sample_designs_co.html, and international.ipums.org/international/sample_designs/sample_designs_mx.html.

[15] IPUMS also offers 2005 census records for Mexico, but these lack the birthplace variable BPLMX, which obstructs their use here.

All new regressions reported here incorporate person-level sampling weights provided by IPUMS. Most U.S. and Colombia IPUMS samples are "flat," meaning that this weighting is not needed to make them statistically representative. However, there are exceptions (Ruggles et al. 2015; usa.ipums.org/usa/intro.shtml#weights). And more of the Brazil and Mexico samples require weighting because of systematic under- and over-sampling of various subpopulations.[16] Bleakley (2010) does not mention using sampling weights. The paper appears to use them in aggregating the outcome variables into birthplace–birth year–census year cells (to form the Y_{ctj}), for I obtain the best matches to the public Bleakley (2010) data when also doing so. However, after aggregation, the Bleakley (2010) regressions are weighted only by the square root of cell size—again, going by what produces the best match—which is based on the unweighted observation counts within cells. Thus, the Bleakley (2010) regressions appear not to fully correct for non-representative sampling within the IPUMS data sets.

My use of IPUMS weights is not pre-registered. However, it is implicitly preregistered in that Roodman (2018) does the same.

4 Reconstruction of analysis data

From IPUMS microdata, I reconstruct all the Bleakley (2010) outcome variables. As for the independent variables, I import reconstructed versions for the United States from the Roodman (2018) replication of Bleakley (2007). I do not attempt to reconstruct the independent variables for the Latin countries, viewing the time cost as prohibitive.[17] In the regressions, I use reconstructed dependent variables where available and take them from the public Bleakley (2010) data otherwise.

To check for problems in the reconstructed variables—and the originals—I compare the two to the degree possible. The public Bleakley (2010) data report the variables in two forms. Long-difference cross-sections contain one observation, in differences, for each geographic unit. Panel data sets aggregate more finely, within birth year–birthplace–census year cells; but they only cover one outcome per country.

Table 2 presents means and standard deviations for all Bleakley (2010) outcomes in the original and new data sets, as well as the cross–data set correlations. All statistics incorporate IPUMS sampling weights. The matches are mostly good, especially in the data arrayed for panel analysis, which is the framework of exclusive interest here. By chance, the panel cross–data set correlations round to 0.931 for the United States and Colombia; the correlation is 1.000 for and 0.998 for Mexico (right side of Table 2).[18,19] In the long-difference data (left side of the table) the correlations

[16] See international.ipums.org/international-action/sample_details.

[17] This paper began as an offshoot of a project to review the evidence of the long-term impact of deworming. Having fully reconstructed the U.S.-focused Bleakley (2007), and discovered the publicly available analysis data for Bleakley (2010) the choices made here amount to picking low-hanging fruit. The only additional variable reconstruction was for the outcomes in the Latin countries, which was made practical by the accessibility of IPUMS International online data system.

[18] Total income in the 1960 Brazil data is reported after censoring into an ordinal variable. Bleakley (2010) appears to "top-code" the 50,000-and-above category as 50,001, so I do the same. For lower categories, range midpoints are used, as documented in the original.

[19] The match with Colombia is most hard-won. After much trial and error, I determined that the "bplcol2" fields of the Columbia data sets, which index the geographic unit, the *municipio*, had been rearranged relative to other variables, as if

are a bit lower for the U.S. outcomes, at around 0.9 and 0.8, and are much lower for earned income in Brazil, at 0.15.[20] Lacking full access to the original data and code, it is hard to know what causes these discrepancies.

In the case of the United States, I copy from Roodman (2018) the reconstructed independent variables. Table 3 does for these variables what Table 2 did for the dependent ones. The first three rows show nearly perfect cross-state matches for the indicator of regional malaria burden (M) as well as the two controls included in all Bleakley (2010) panel specifications, a state-level measure of agricultural wages in 1899 and a dummy for being in the South. The remaining rows turn to the variables introduced in Bleakley (2010)'s "full controls" specifications, which are the focus here.[21] The matches are close, except in the education variables. These mismatches are unsurprising given the ambiguity in the Bleakley descriptions of the education variables, which are defined as changes during the date range "circa 1902–32." Most likely the reconstructed variables use different editions of the underlying federal government report. And possibly the negative correlation for log change in pupils per teacher owes to Bleakley (2010) inverting this variable, to teachers per pupil—which itself would be harmless when controlling for log changes.

The juxtaposition of original and reconstruction also exposes discrepancies between the Bleakley (2010) text and the Bleakley (2010) data, some of which appear to be implementation errors. Since the publicly available data and code exactly replicate the published Bleakley (2010) tables, the published results reflect all these departures from the text. In particular, the cross-state control variables are to have been multiplied by Exp before entering the regressions; they are multiplied by birth year instead. While the text defines Exp assuming childhood lasts 21 years, in the panel data, Exp in fact rises from 0 to 1 over 18 years. (Likewise for Brazil, Colombia, and Mexico.) The control "Doctors per Capita, 1998" is actually residents per doctor. The main text lists the log change in teacher salaries among the controls but Bleakley (2010) Appendix III and the code refer instead to the log change in school term length. The U.S. panel regressions include birth cohorts back to 1815, which is earlier than the 1825 starting point stated in text.

Table 4, below, checks whether these problems drive the Bleakley (2010) U.S. panel results. The table closely follows the format of Table 4, panel A, of Bleakley (2010), which presents all the U.S. panel estimates, except that it doubles the number of columns. The odd columns copy from the original. The even columns present results obtained from the public Bleakley (2010) data set after fixing the apparent errors: properly constructing the interaction terms with Exp instead of birth year, inverting residents per doctor to doctors per resident, and defining childhood as lasting 21 years. As well, observations are weighted by cell size rather than the square root thereof, as set forth in the pre-analysis plan. These fixes (largely not pre-registered) cause no substantive change

the column had been sorted in Excel while leaving other columns untouched. Thus, the variable does not in fact obey the coding of the IPUMS International field from which it ultimately derives, BPLCO2. After consulting the primary source for the altitude and temperature variables (Banco de la Republica 1960), I estimate that the mapping to IPUMS codes can be recovered from the Bleakley (2010) public long-difference data using the following algorithm. Sort the data set by bplcol1 and bplcol2; then number the rows starting from 1, except skipping indexes 284 and 473. I cannot tell whether only bplcol2 was rearranged relative to the rest of the data set—which in itself would not affect the Bleakley (2010) results—or whether other variables were too, which would be an error.

[20] For Brazil, total income, as distinct from earned income, is of primary interest in the analysis, partly because more census rounds collected it, partly because it matters more.

[21] Bleakley (2010) Figure 4 is the sole figure in the original exploring the temporal evolution of the β_t in (4). Its specifications all include the full control set.

in the Bleakley (2010) panel results.

5 Time series results

Having reconstructed all of the Bleakley (2010) variables except for the cross-sectional independent variables for Brazil, Colombia, and Mexico, I implement the revised designs defined in section 2. To start, Figure 1, below, strives to imitate Bleakley (2010) Figure 4, the sole presentation in the original of time series results. Each data point represents an estimate based on (3) of β_t, which is the cross-sectional association, among people born in year t, between historical malaria burden in place of birth and adult earnings. The graph uses only public Bleakley (2010) data, which aggregates from samples of (white) men. For the United States, the dependent variable is log occupational income score; for Brazil and Mexico, log total income; and for Colombia, the log of a Bleakley (2010)-constructed variable called the industrial income score. Each birth cohort–specific regression includes mean-reversion controls, dummies for national regions, and additional controls, all of which vary in definition by country, as laid out in the Bleakley (2010) appendix. The new figure departs substantively from the original only in drawing 95 percent confidence intervals for the point estimates. It departs cosmetically in not superimposing a plot of the Exp step function. But vertical lines are drawn to mark the birth cohorts at which Exp kinks—the years the eradication campaigns began, and 18 years before (since 18 is used in the original data and code, rather than the 21 stated in the original text).

Figure 1 matches Bleakley (2010) Figure 4 well, but not perfectly. This is to be expected when original data is used, but original code is not. (Recall that the public Bleakley (2010) code only generates tables, not figures.) In all four countries, β_t rises with time—generally from negative values toward zero, but in Colombia from approximately zero to positive values.

Figure 2 updates Figure 1 by fitting to the expanded data sets at the microdata level, according to (4). Now, census samples are added or increased in density. Women are included. For the United States, blacks are added too. In tandem, sex and (for the U.S.) race dummies enter the control set, fully interacted with the age/census round dummies. Observations are weighted using IPUMS individual weights. Standard errors are clustered across birth cohorts, by the state or *municipio* of birth. In marking the first potential kink point, childhood is taken to last 21 years, as stated in the Bleakley (2010) text.

Except in Mexico, the expanded-sample results appear statistically compatible with the smaller-sample results. In Mexico, an apparent rise before the predicted take-off year of 1936 now disappears, improving the match with the Bleakley (2010) prediction.

Figure 2 confronts us with the paramount empirical question in this reanalysis: did the cross-sectional association between baseline malaria endemicity and future earnings rise at an historically anomalous rate among the cohorts born in the run-up to eradication, the period demarcated by the dashed, vertical grey lines? A glance at Figure 2 tentatively suggests that the answer is "yes" in all the countries save Mexico.

To formally test that interpretation, Figure 3, Figure 4, and Figure 5 fit the polynomial, fixed-spline, and flexible-spline models, defined in section 2, to the expanded microdata. These figures

retain the dots from Figure 2 but, for legibility, drop the confidence intervals.

The polynomial fits, in Figure 3, largely support the Bleakley (2010) impact model. The fits for models of order 0 to 5 are shown in orange, green, blue, red, purple, and brown, respectively, while the six corresponding p-values for the coefficient on $Exp \times M$ are listed beneath. The estimates of the coefficient on $Exp \times M$ in (5) are gathered in Table 5. For all four countries—for the income proxies used in Bleakley (2010) Figure 4—the BIC-minimizing polynomial orders assign a statistically strong positive value to β, albeit less so for Mexico ($\beta = 0.133$; standard error $= 0.069$). The BIC-favored orders are 2 for the U.S. and Colombia, 0 for Brazil, and 1 for Mexico.

The fixed-spline fits to the income proxies, in Figure 4, tell a similar story. The hypothesis of no acceleration at the first kink is comfortably rejected in Latin America ($p = 0.00, 0.00, 0.07$ for Brazil, Colombia, and Mexico). An upward bend in the United States appears to have begun earlier than predicted in the Bleakley (2010) impact model, making the null of no slope change at the predicted time much harder to reject ($p = 0.39$). Meanwhile, the null of no slope change at campaign onset (second kink point) is strongly rejected for the United States, Brazil, and Mexico ($p = 0.03, 0.00, 0.05$) but less so for Colombia ($p = 0.23$). These results broadly corroborate Bleakley's step-like form for the impact of malaria eradication across cohorts. The weak results for the U.S. might reflect the historical reality that malaria eradication efforts came in several waves in the 1920s, 1930s, and 1940s (Barreca, Fishback, and Kantor 2012), whereas the efforts in Latin American were more compressed in time.

Turning to the flexible-spline fits in Figure 5, most of the data-chosen kink dates line up reasonably well with those chosen *a priori* by Bleakley (2010). The fit is especially close for Brazil, where the fit kinks upward at 1934 and downward at 1960 instead of the predicted 1936 and 1957. The fit is weaker in the U.S., where the ramp-up runs from 1887 to 1932 instead of 1899 to 1920; and in Mexico, where the second kink appears in 1947, ten years before the predicted 1957.

Last, Figure 6, Figure 7, and Figure 8 apply the methods of the previous three figures to the outcomes for which Bleakley (2010) reports long-difference but not panel results. These are Duncan's socioeconomic indicator (SEI) for the United States, earned income for Brazil, and literacy and years of schooling for all three Latin countries.

Somewhat like the Bleakley (2010) long-difference regressions, the new figures produce a more mixed bag for these outcomes. For the polynomial models, Figure 6 displays the plots and Table 5 the corresponding impact estimates and standard errors. Forced to fit to the full U.S. historical record, the polynomial models rather confidently endow the treatment term $Exp \times M$ with explanatory power. Polynomial controls also strengthen the fit for earned income in Brazil. For human capital variables, signs, magnitudes, and statistical significance of the impact estimates vary substantially with the polynomial order, which is easier to see in Table 5. Suggestions of impact do not appear robust.

Figure 7 presents the fixed-spline fits for these additional outcomes. In the United States, the trend on Duncan's SEI appears to bend at the first allowed kink, but not at all at the second, reversing the pattern for the closely related occupational income score (refer back to the upper-left of Figure 4). In Brazil, while relative progress on earned income (as distinct from total income) slows

when expected, it does not appear to accelerate when expected (21 years earlier), perhaps owing to low statistical power from small samples in the early years. In none of the Latin countries does relative progress on adult literacy or years of schooling slow as predicted at the second kink point. In all, the trend bends with statistical significance at the first kink point—but bends the "wrong" way in Mexico, downward.

Moving to the flexible-spline fits of these outcomes, in Figure 8, the main update is that the relative rises in Brazil in literacy and years of schooling look more attributable to the malaria eradication campaign, if we allow some give in the kink dates. For instance, the plot for years of schooling (second row, on left) suggests a sharp acceleration around 1942 and deceleration in 1961. 1961 is later than Bleakley's preferred campaign start year of 1957, but the latter is an abstraction from a complex reality: campaigns lasted years and began at different times in different countries, and in different regions of the same country.

Overall, the new time series results support the proposition that reduced childhood malaria exposure increased adult earnings in Latin America. It may well have done so in the United States too, but there the spline fits less consistently point to acceleration and deceleration with the expected timing (top left of Figure 4 and of Figure 7). Eradication did not so clearly and consistently increase schooling or literacy.

To test whether my expansion of the samples to women and blacks (the latter in the U.S. only) affects these findings, the appendix repeats the analysis embodied in the last six figures while restoring Bleakley (2010)'s restriction to (white) men. The pattern of results changes little. Perhaps the most significant change is that the β_t contour for the U.S. socioeconomic index now better fits the Bleakley (2010) step-shaped impact model. In the fixed-spline model the p-value for an upward trend break at the expected time falls from 0.39 (upper left of Figure 4) to 0.08 (upper left of Figure A2). And the dates chosen by the flexible-spline model, 1894 and 1921, closely correspond to the 1899 and 1920 predicted by Bleakley (2010), as shown in the upper left of Figure A3. On the other hand, results for the closely related Duncan's SEI remain much weaker (e.g., upper left of Figure A5 and Figure A6).

6 Conclusion

Bleakley (2010) identifies impacts from variation in the product of two factors: the geographic pattern of baseline malaria burden and the timing of campaigns to relieve that burden. The first factor cannot credibly be viewed as exogenous since it is a marker for climate and geography, and thus local economic history. The second can be taken as exogenous, but only in the short term. That malaria eradication campaigns took place between, say, 1900 and 2000, is of a piece with the economic and scientific development of the Americas and to that extent endogenous to broader forces that could also affect earnings and human capital accumulation. That individual eradication efforts started precisely when they did, rather than a few years sooner or later, is more an accident of history. Thus, given the informal priors I bring to this study, for it to produce strong evidence of impact, its results must match predictions of certain trend breaks, and that with an accuracy measured in years, not decades.

In my view, only the time-series analysis performed here fully confronts the challenge of generating evidence to test these predictions with the requisite precision. The Bleakley (2010) long-difference regressions speak to whether relative gains occurred in historically malarial areas but not to the relationship of this convergence to longer-term trends. The Bleakley (2010) panel regressions get more at functional form, introducing birthplace-specific quadratic time controls. But as presented, it is hard to judge whether these models are specified flexibly enough to largely absorb ambient trends. By graphing the time series patterns and performing formal inference on them, the present paper provides a clearer view of the temporal variation that is the most credible source of causal identification.

The reanalysis does not trigger much update. Bleakley (2010) finds "that cohorts with less childhood exposure to malaria have higher literacy rates, but results are mixed for years of schooling." The new analysis tends to produce mixed results for both. Meanwhile, it broadly supports Bleakley's "main result" that the evidence indicates that eradication raised adult income. That it does so more clearly for Latin America than the United States might owe to the sharper onset of the Latin campaigns. This largely supportive conclusion contrasts with that from the separate reanalysis of Bleakley (2007)'s assessment of *hookworm* eradication in the American South 100 years ago (Roodman 2018).

Separately, this reanalysis points up limitations in the data and code archiving practices of the American Economic Association journals. One purpose of those archives is to increase confidence in published results by documenting precisely how they are obtained. Current archiving practices undercut this purpose in two respects. First, they provide no access to the primary data, or at least to the code that transforms the primary data into the analysis data. The *American Economic Review's* own assessment of compliance with its data availability policy highlighted this omission in 2011. "Simply requiring authors to submit their data prior to publication may not be sufficient to improve accuracy. . . .The broken link in the replication process usually lies in the procedures used to transform raw data into estimation data and to perform the statistical analysis, rather than in the data themselves" (Glandon 2011). Second, code is provided for tables only, not figures. Yet figures can play a central role in a study's conclusions and impact. Like tables, figures distill large amounts of data to inform inference. They ought to be fully replicable, but only can be if their code is public too.

As a result of these two gaps, to the extent that Bleakley (2010) and this reanalysis directly contradict one another, it is impossible to be sure why. And to the extent they agree when the reanalysis copies variables from the publicly archived data, one cannot know to what extent the shared conclusions are driven by bugs in the (non-public) transformation code. These avoidable ambiguities mis-serve the researchers and decisionmakers that journal authors and publishers aspire to influence.

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.

Banco de la Republica (1960). "Atlas de Economia Colombiana." Bogota, Colombia: Departamento de Imprenta del Banco de la Republica. Available at ufdc.ufl.edu/uf00077026/00002.

Barreca, Alan, Price V. Fishback, and Shawn Kantor (2012). "Agricultural policy, migration, and malaria in the United States in the 1930s." *Explorations in Economic History* 49 (4): 381–98. DOI: 10.1016/j.eeh.2012.05.003.

Bleakley, Hoyt (2006). "Malaria in the Americas: A retrospective analysis of childhood exposure." Working Paper 2006-35. Centro de Estudios sobre Desarrollo Económico, Universidad de los Andes. Bogotá. Available at sticerd.lse.ac.uk/seminarpapers/dg12032007.pdf.

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 (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.

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.

Glandon, Philip J. (2011). "Appendix to the Report of the Editor: Report on the American Economic Review Data Availability Compliance Project." *American Economic Review* 101(3), 696–699. DOI: 10.1257/aer.101.3.684.

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.

Minnesota Population Center (2017). *Integrated Public Use Microdata Series, International: Version 6.5*. Minneapolis: University of Minnesota. DOI: 10.18128/D020.V6.5.

Roodman, David (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: 110.18718/81781.7.

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

Table 1: U.S. IPUMS census sample densities in original and expanded data sets

Census year	Original (percent)	Expanded (percent)
1860	0	1.2[a]
1870	0	1.2
1880	100	100
1890	0	0
1900	1	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	5	5

[a]Excludes slaves.

Table 2: Summary statistics of Bleakley (2010) dependent variables

	Long-difference cross-section			Panel data		
	Original	New	Correlation	Original	New	Correlation
United States						
Log occupational income score	0.324 (0.084)	0.292 (0.082)	0.897	3.286 (0.112)	3.279 (0.132)	0.931
Log Duncan's SEI	0.560 (0.102)	0.504 (0.074)	0.806			
Observations	48	48		9604	9605	
Brazil						
Log total income	–0.012 (0.080)	–0.007 (0.087)	0.945	8.625 (2.268)	8.701 (2.262)	1.000
Log earned income	–0.012 (0.075)	–0.017 (0.223)	0.152			
Literacy	–0.002 (0.053)	–0.003 (0.052)	0.994			
Years of schooling	0.032 (0.537)	–0.001 (0.513)	0.949			
Observations	24	28		2156	2453	
Colombia						
Industrial income score	–0.049 (0.080)	–0.055 (0.098)	0.855	–0.112 (0.174)	–0.106 (0.200)	0.931
Literacy	–0.020 (0.100)	–0.018 (0.096)	0.973			
Years of schooling	–0.480 (0.632)	–0.487 (0.617)	0.973			
Observations	523	525		38070	39513	
Mexico						
Log earned income	–0.044 (0.173)	–0.110 (0.328)	0.925	9.659 (2.914)	9.391 (2.902)	0.998
Literacy	–0.017 (0.072)	–0.021 (0.082)	0.993			
Years of schooling	–0.229 (0.497)	–0.373 (0.529)	0.931			
Observations	32	32		2965	2965	

Variable means displayed with standard deviations in parentheses beneath. Third and sixth columns show cross–data set correlations. "Original" results computed from public Bleakley (2010) data. "New" results computed after reconstructing the data sets. All statistics weighted by cell-level sums of the IPUMS-provided individual weights in the reconstructed data set.
Source: Author's calculations.

Table 3: Summary statistics of U.S. cross-state variables

Variable	Original	New	Correlation
Malaria share of mortality, 1889 (M)	0.318 (0.326)	0.295 (0.302)	0.994
Agricultural wage, 1899 ($/month)	16.938 (6.393)	17.415 (6.396)	0.999
South	0.271 (0.449)	0.271 (0.449)	1.000
Residents per doctor, 1898	743.333 (244.706)	743.361 (244.719)	1.000
Board of health spending, 1898 ($/1,000 residents)	6.333 (13.253)	6.779 (13.321)	0.976
Infant mortality rate, 1890 (per 1,000 births)	162.797 (68.310)	105.358 (51.474)	0.983
Hookworm prevalence among army recruits, 1917–19	0.069 (0.097)	0.069 (0.097)	1.000
Log change in teacher salaries, circa 1902–32	1.444 (0.175)	3.216 (0.199)	0.775
Log change in school term length, circa 1902–32	0.114 (0.122)	0.169 (0.149)	0.631
Log change in pupils/teacher, circa 1902–32	0.118 (0.275)	–0.043 (0.172)	–0.362
Adult literacy rate, 1910	0.907 (0.074)	0.907 (0.074)	1.000
Population urban, 1910	0.340 (0.231)	0.392 (0.225)	0.982
Population black, 1910	0.107 (0.164)	0.107 (0.163)	1.000
Male unemployment, 1930	0.043 (0.018)	0.079 (0.026)	0.913
Observations	48	48	

Variable means displayed with standard deviations in parentheses beneath. Final column shows cross–data set correlations. All statistics are unweighted. "Original" results computed from public Bleakley (2010) data. "New" results computed after reconstructing the data set from primary sources. Sample excludes Alaska, Hawaii, and the District of Columbia.
Source: Author's calculations.

Table 4: Replication of Bleakley (2010) Panel estimates of the effect of childhood exposure on log occupational income score in the United States

Degree of polynomial trend for year of birth	Mean reversion and region controls						Additional controls					
	0		1		2		0		1		2	
	Original	New	Original	New	Original	New	Original	New	Original	New	Original	New
Baseline	0.131 (0.030)	0.183 (0.038)	0.115 (0.031)	0.196 (0.038)	0.131 (0.025)	0.093 (0.024)	0.120 (0.024)	0.174 (0.036)	0.098 (0.035)	0.199 (0.044)	0.116 (0.027)	0.055 (0.012)
Post-1920 break in birthplace time trend	0.082 (0.015)	0.103 (0.016)	0.094 (0.020)	0.139 (0.023)	0.105 (0.024)	0.073 (0.017)	0.073 (0.020)	0.100 (0.016)	0.080 (0.020)	0.140 (0.027)	0.085 (0.021)	0.056 (0.013)
Allow for birth-place × time effects	0.103 (0.026)	0.108 (0.016)	0.110 (0.030)	0.138 (0.021)	0.123 (0.023)	0.079 (0.017)	0.089 (0.030)	0.106 (0.017)	0.092 (0.033)	0.138 (0.025)	0.108 (0.025)	0.066 (0.014)
Drop early census years (<1930)	0.106 (0.021)	0.107 (0.016)	0.105 (0.017)	0.084 (0.018)	0.032 (0.015)	0.014 (0.014)	0.096 (0.014)	0.107 (0.016)	0.109 (0.023)	0.068 (0.015)	0.032 (0.019)	0.014 (0.014)
Add region×year×YOB effects	0.131 (0.030)	0.175 (0.038)	0.116 (0.029)	0.194 (0.037)	0.131 (0.024)	0.090 (0.025)	0.123 (0.025)	0.166 (0.036)	0.102 (0.034)	0.197 (0.043)	0.119 (0.027)	0.050 (0.013)

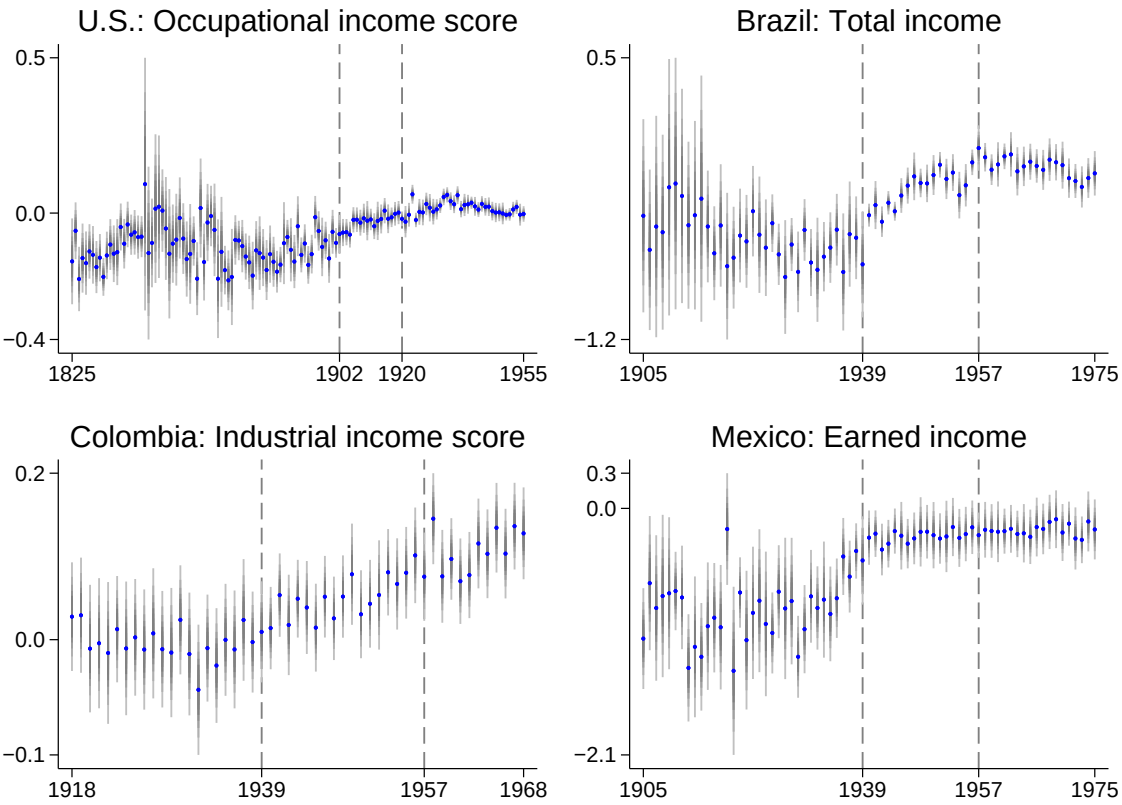
Each cell reports an OLS estimate of the association between log occupational income score and the product of the potential childhood malaria exposure variable (Exp) and pre-campaign malaria intensity (M), according to (2). Fixed effects are allowed for each census year, each birth year, and each state. Geographic controls, for which results are not shown, enter the specification interacted with Exp. All variables are described in the Bleakley (2010) appendix. Before estimation, variables are averaged into birth year–census year–state of birth cells. The columns marked "mean reversion and region controls" use the basic controls sets, while the "additional controls" columns use the "full controls" specifications. Within each group, the degree of the polynomial time trend control is varied column-wise and the control set and sample are varied row-wise, as explained in Bleakley (2010). "Original" results are generated with Bleakley (2010) public data and code and exactly match the original. "New" results use the same data and address coding issues described in text. Standard errors, shown in parentheses, are clustered on state of birth. Observations are weighted by the square root of the cell size in the original regressions and by cell size in the new regressions.

Table 5: Impact estimates on all Bleakley (2010) outcomes, controlling for polynomial time trend up to order 5

Country	Outcome	Coefficient on $M \times Exp$					
	Order of Polynomial Trend	0	1	2	3	4	5
U.S.	Occupational income score	0.087	0.070	**0.064**	0.041	0.040	0.015
		(0.010)	(0.012)	**(0.014)**	(0.015)	(0.014)	(0.016)
	BIC	–17.32	–32.27	**–46.18**	–31.48	–33.66	–35.38
	Duncan's SEI	0.096	0.068	0.056	0.042	**0.032**	0.031
		(0.023)	(0.035)	(0.033)	(0.032)	**(0.033)**	(0.027)
	BIC	169.87	149.63	96.55	114.52	**82.89**	88.33
Brazil	Total income	**0.439**	0.477	0.486	0.504	0.433	0.253
		(0.071)	(0.132)	(0.120)	(0.082)	(0.097)	(0.122)
	BIC	**105.62**	110.12	113.22	117.74	117.58	120.00
	Earned income	**0.276**	0.285	0.340	0.323	1.000	0.692
		(0.060)	(0.134)	(0.113)	(0.103)	(0.180)	(0.212)
	BIC	**83.21**	87.20	91.45	95.58	92.78	91.75
	Literacy	0.121	0.009	0.048	–0.040	–0.102	**0.069**
		(0.026)	(0.037)	(0.032)	(0.033)	(0.041)	**(0.046)**
	BIC	48.33	32.55	–2.76	–1.96	–30.88	**–33.26**
	Years of schooling	0.846	0.883	0.870	0.906	**0.256**	0.906
		(0.358)	(0.565)	(0.596)	(0.458)	**(0.790)**	(0.729)
	BIC	275.74	280.18	274.92	279.29	**265.76**	269.75
Colombia	Industrial income score	0.031	0.018	**0.039**	0.025	0.029	0.170
		(0.009)	(0.011)	**(0.012)**	(0.021)	(0.023)	(0.058)
	BIC	–127.23	–124.15	**–147.42**	–143.86	–139.91	–144.93
	Literacy	0.020	0.009	0.018	**–0.020**	–0.006	–0.011
		(0.012)	(0.010)	(0.010)	**(0.018)**	(0.019)	(0.035)
	BIC	–166.30	–165.05	–167.32	**–170.26**	–168.59	–164.50
	Years of schooling	0.368	–0.015	**0.180**	0.151	0.303	–0.079
		(0.156)	(0.176)	**(0.151)**	(0.382)	(0.352)	(0.758)
	BIC	160.49	154.20	**138.83**	142.90	145.20	149.02
Mexico	Earned income	0.250	**0.133**	0.199	0.255	0.274	–0.146
		(0.051)	**(0.069)**	(0.062)	(0.136)	(0.124)	(0.278)
	BIC	154.25	**152.90**	156.53	160.90	163.42	166.48
	Literacy	0.015	–0.031	–0.052	**0.012**	0.019	0.138
		(0.030)	(0.030)	(0.025)	**(0.021)**	(0.023)	(0.051)
	BIC	–98.93	–96.11	–120.56	**–129.36**	–127.21	–127.13
	Years of schooling	–0.386	–0.433	–0.511	0.542	**0.895**	1.266
		(0.276)	(0.424)	(0.351)	(0.403)	**(0.439)**	(0.605)
	BIC	275.56	280.15	279.17	269.40	**258.19**	262.33

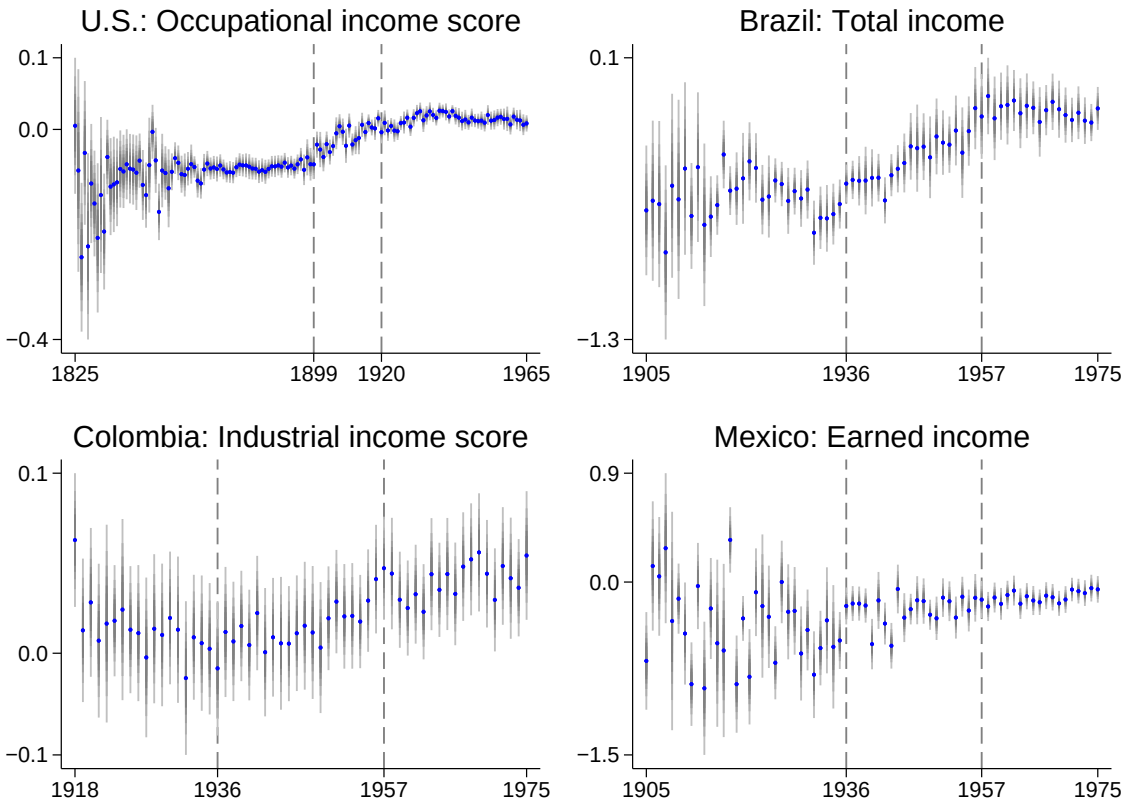
Estimates based on expanded data set, including women and, in the U.S. case, blacks as well as whites. Regressions weighted by IPUMS-provided sampling weights. Standard errors clustered by state of birth in parentheses. BIC is the Bayesian Information Criterion, taking sample size as the number of birth cohorts in each sample and mean-squared error from data points and model fits presented in Figure 3. Bolded results in each row are those favored by the BIC.
Source: Author's calculations.

Figure 1: Replication and extension of Bleakley (2010) Figure 4: Original data sets



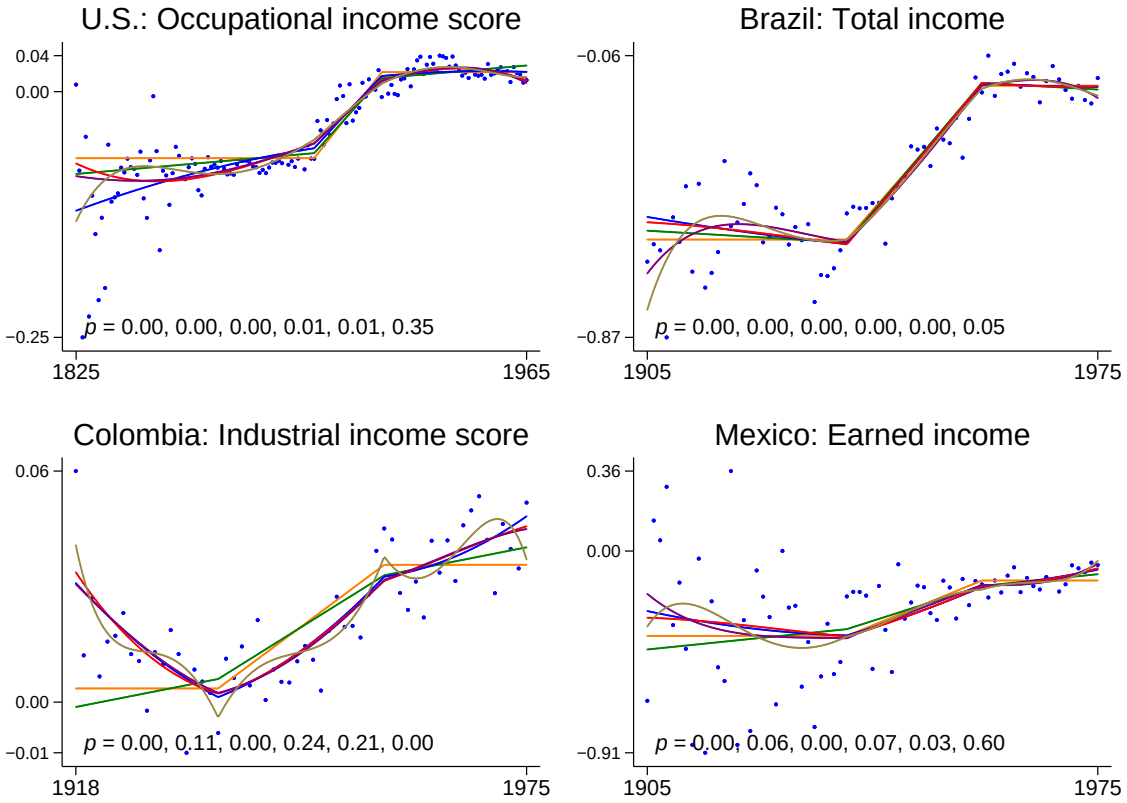
Notes: Blue dots depict point estimates of the cross-geography association of baseline malaria prevalence with the outcome shown within each birth cohort. Grey bars show 95% confidence intervals. Vertical grey lines indicate kink points in Bleakley (2010) exposure function, Exp, which bends upward at the first and plateaus at the second.

Figure 2: Replication and extension of Bleakley (2010) Figure 4: Expanded data sets



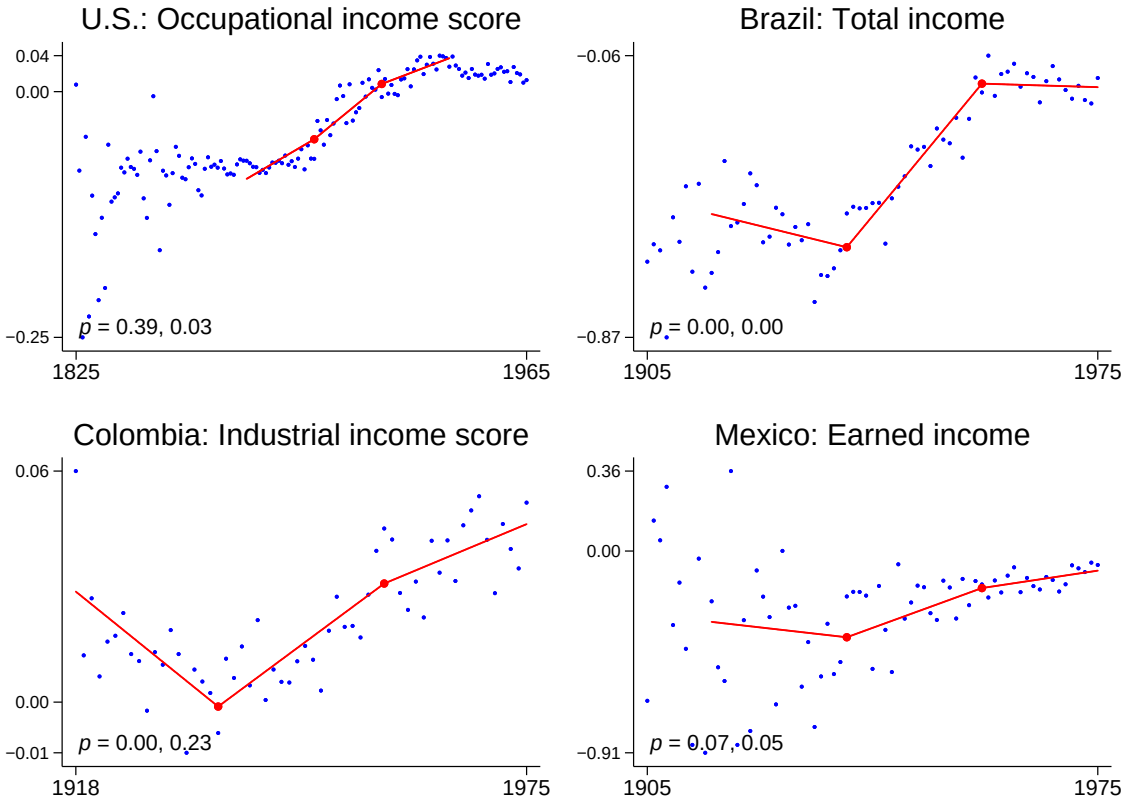
Notes: Blue dots depict point estimates of the cross-geography association of baseline malaria prevalence with the outcome shown within each birth cohort. Grey bars show 95% confidence intervals. Vertical grey lines indicate kink points in Bleakley (2010) exposure function, Exp, which bends upward at the first and plateaus at the second.

Figure 3: Replication and extension of Bleakley (2010) Figure 4: Model with polynomial time controls, fit to expanded data set



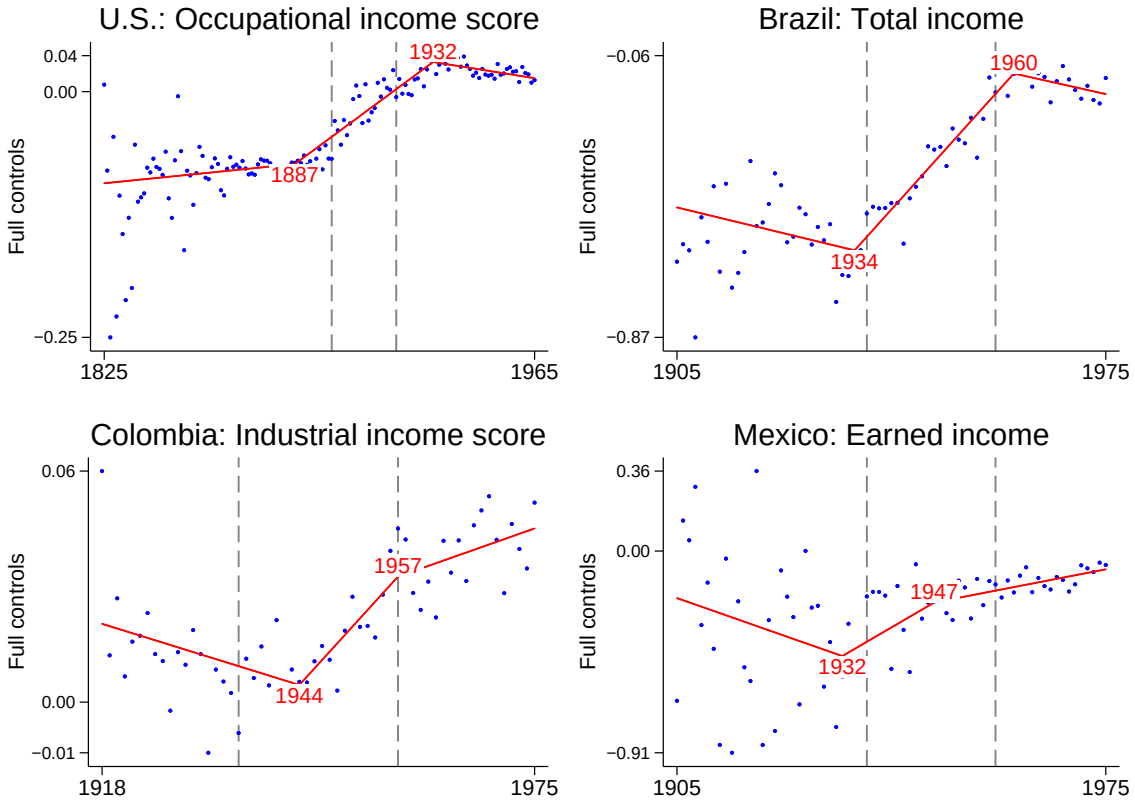
Notes: Dots depict same point estimates as in Figure 2. Each contour represents the best fit of a linear model with the Bleakley (2010) exposure function, Exp, and polynomial time controls ranging in order from 0 to 5. Fits for orders 0–5 are drawn in orange, green, blue, red, purple, and brown, respectively. 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 4: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, fixed kink dates



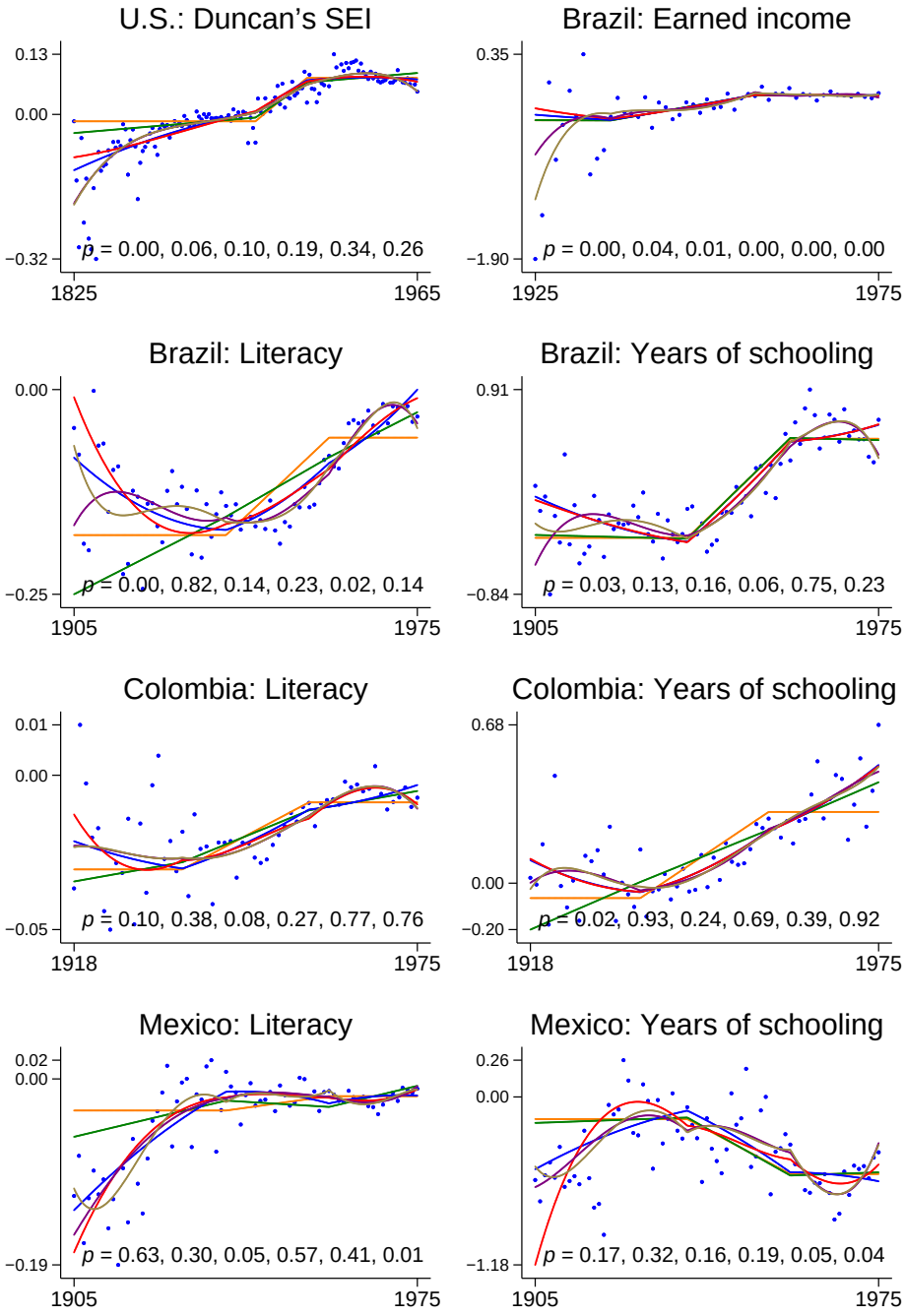
Notes: Blue dots depict same point estimates as in Figure 2. Red contours depict best fits of a piecewise-linear model allowed to kink at the same dates as the Bleakley (2010) exposure function. Each segment spans up to 21 years. p-values in each pane are, respectively, for the nulls of no slope change 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 5: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, flexible kink dates



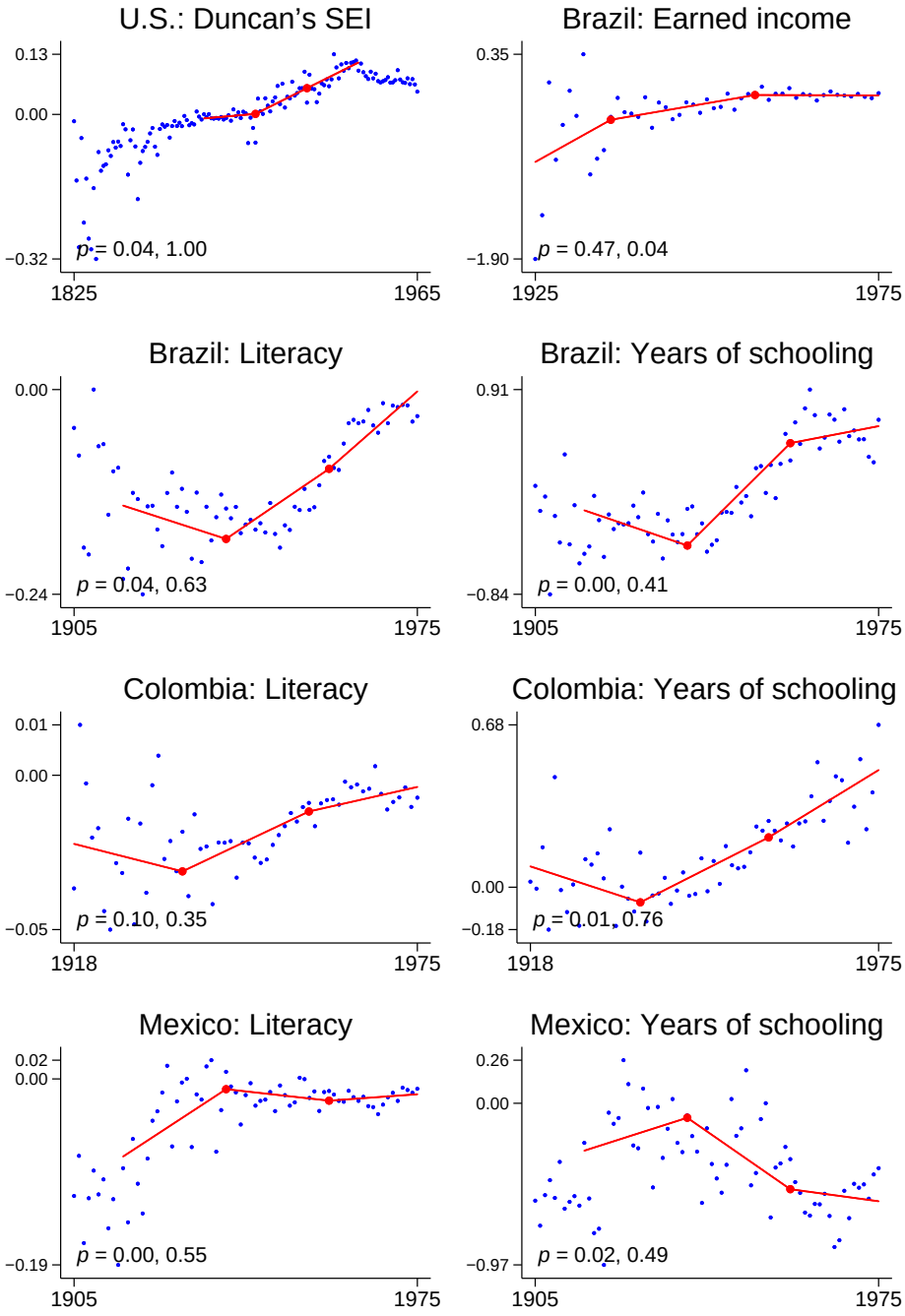
Notes: Blue dots depict same point estimates as in Figure 2. Red contours depict best fits of a piecewise-linear model allowed to kink twice, and fit using the mean-squared-error criterion.

Figure 6: Replication and extension of Bleakley (2010) Figure 4: Model with polynomial time controls, fit to expanded data set, alternative outcome measures



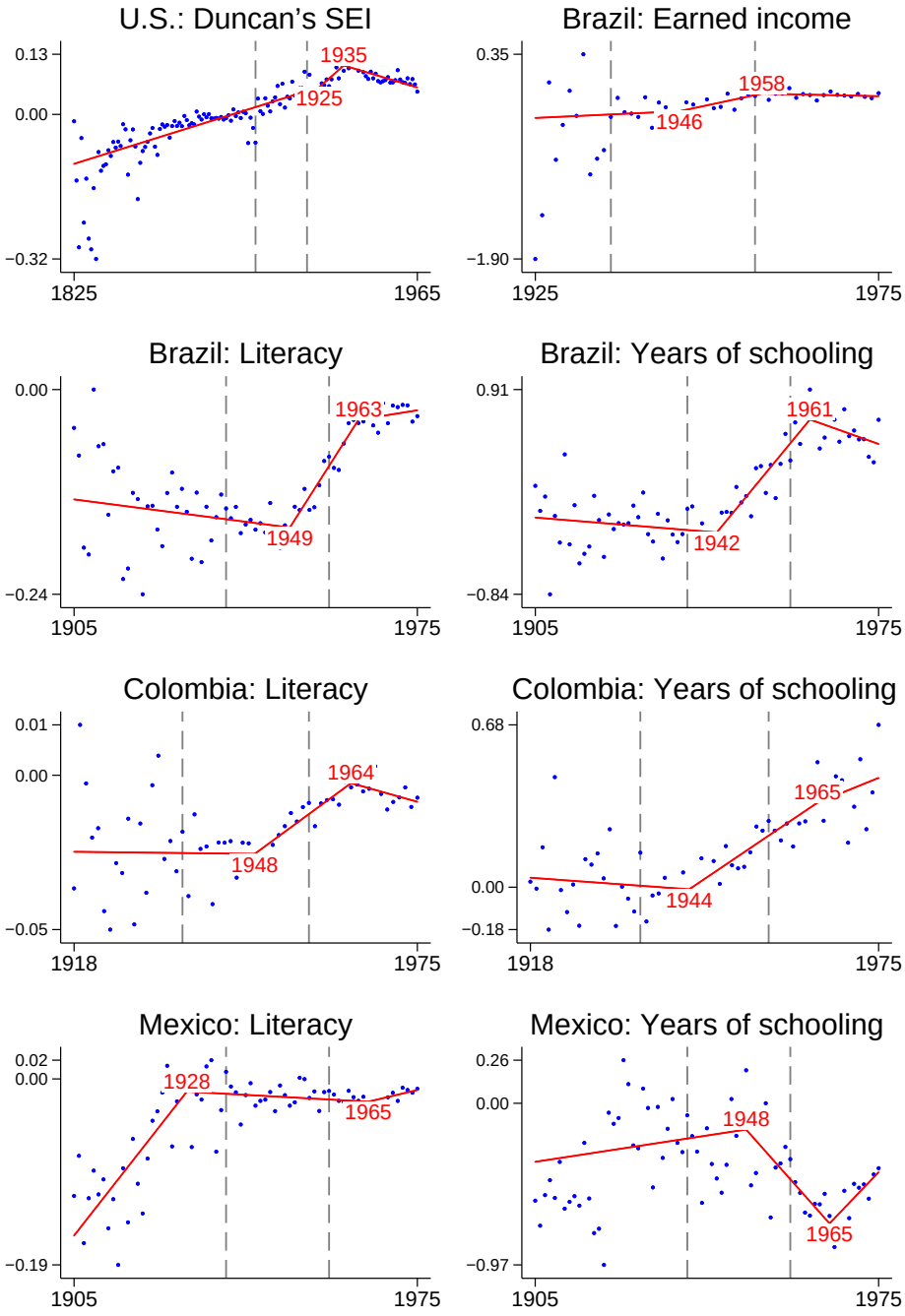
Notes: See notes for Figure 3.

Figure 7: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, fixed kink dates, alternative outcome measures



Notes: See notes for Figure 4.

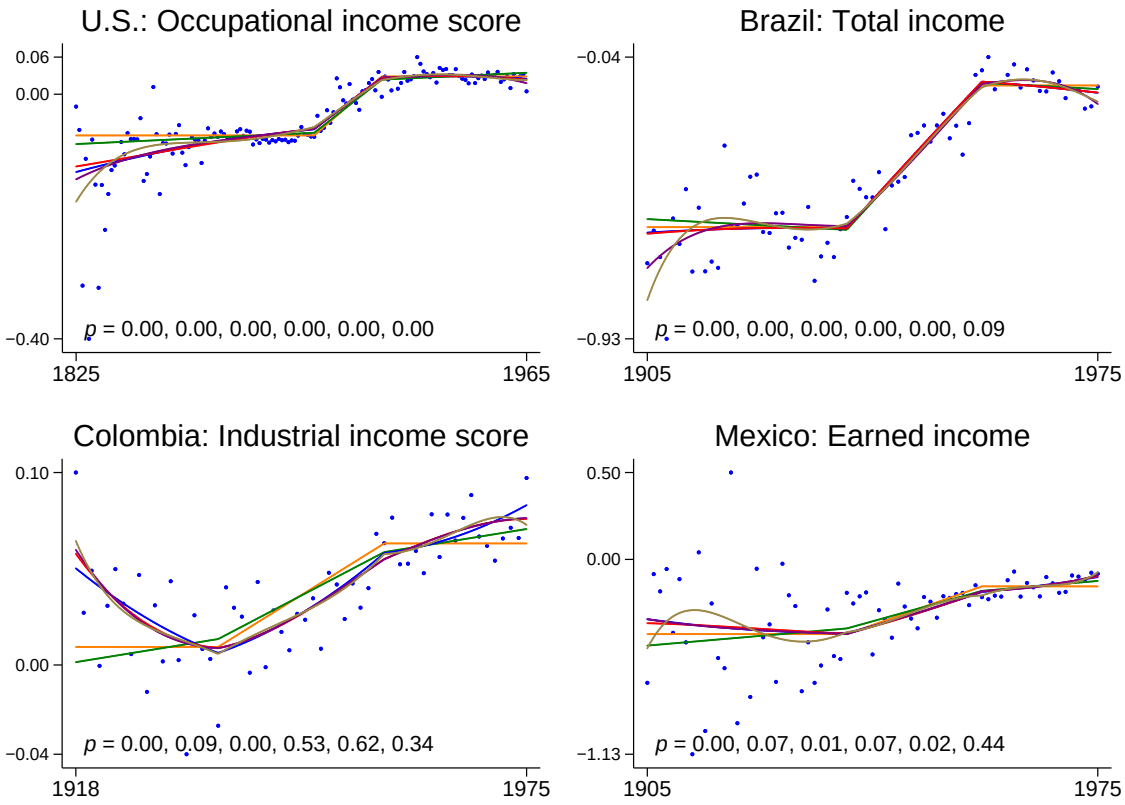
Figure 8: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, flexible kink dates, alternative outcome measures



Notes: See notes for Figure 5.

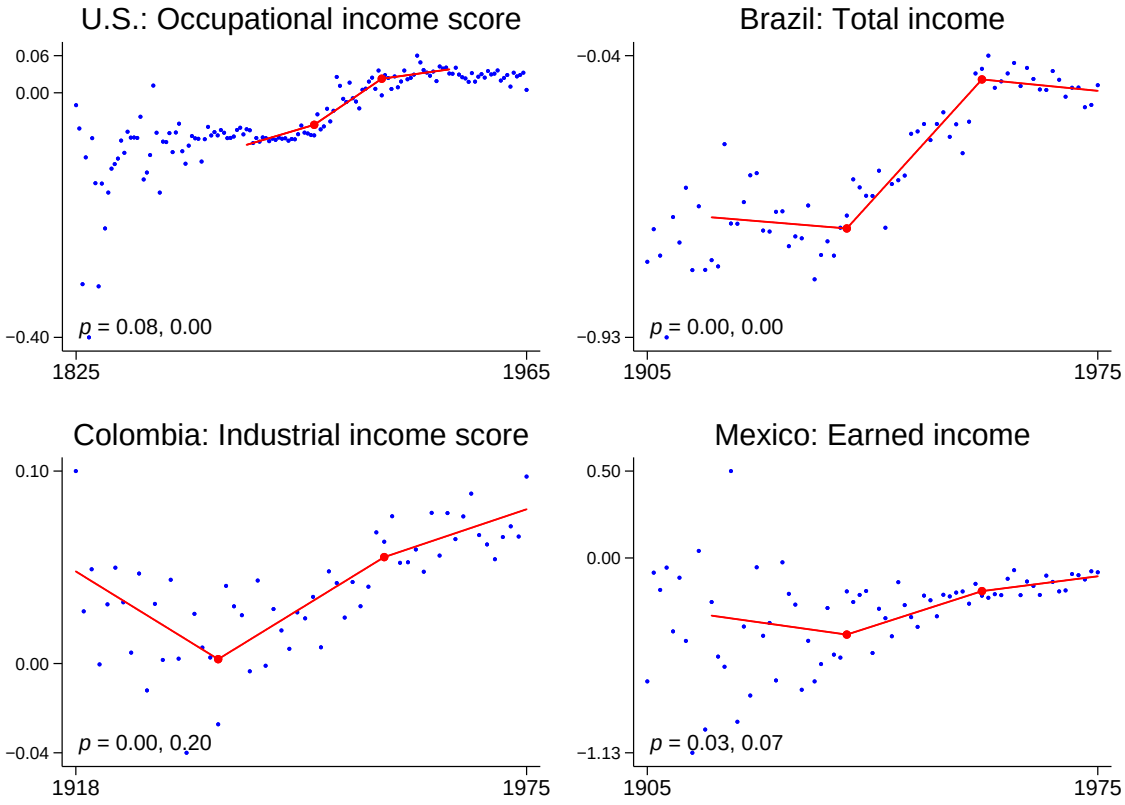
8 Appendix - Additional Figures

Figure A1: Replication and extension of Bleakley (2010) Figure 4: Model with polynomial time controls, fit to expanded data set, excluding women and (in U.S.) blacks



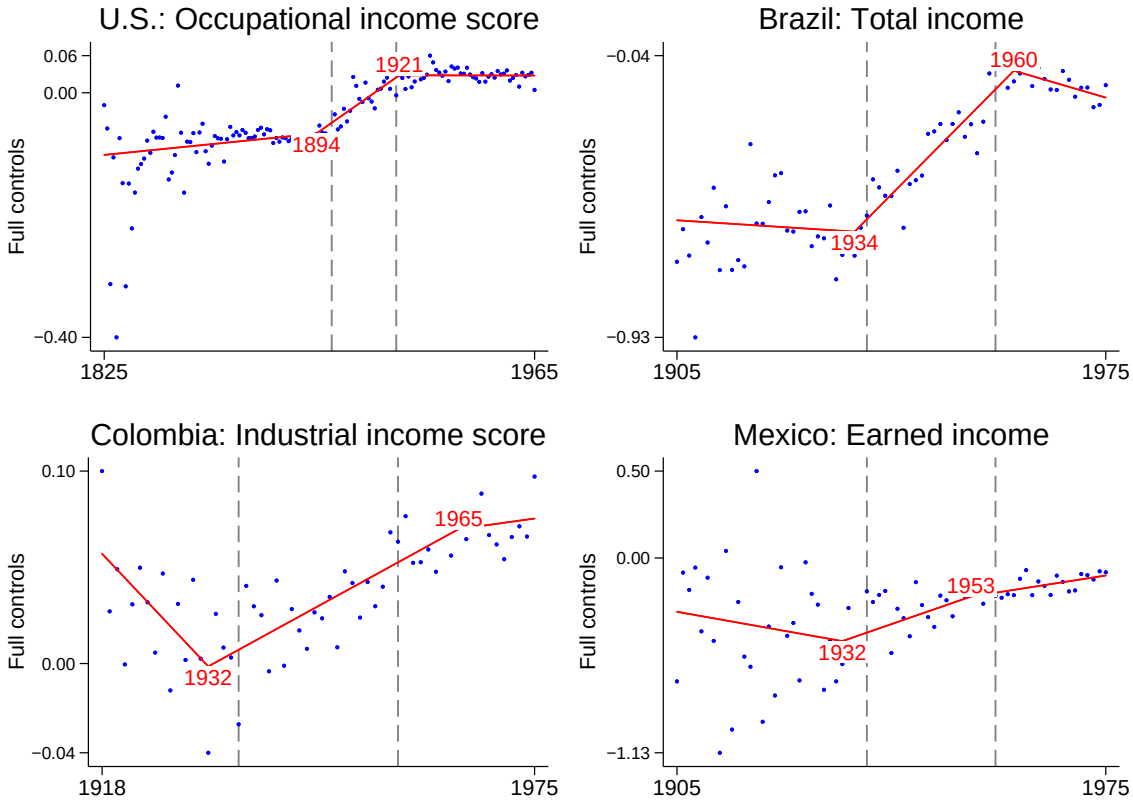
Notes: See notes for Figure 3.

Figure A2: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, fixed kink dates, excluding women and (in U.S.) blacks



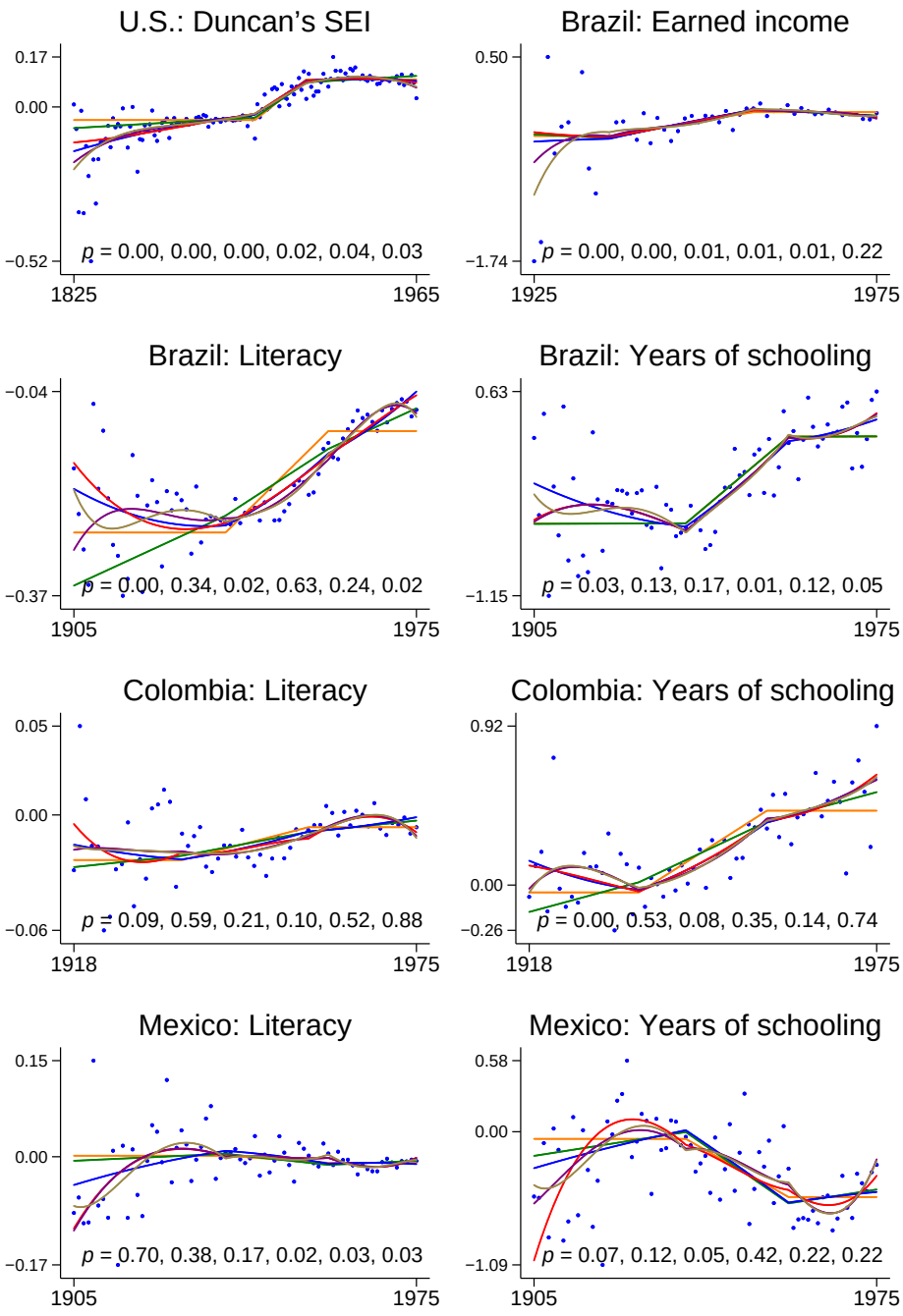
Notes: See notes for Figure 4.

Figure A3: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, flexible kink dates, excluding women and (in U.S.) blacks



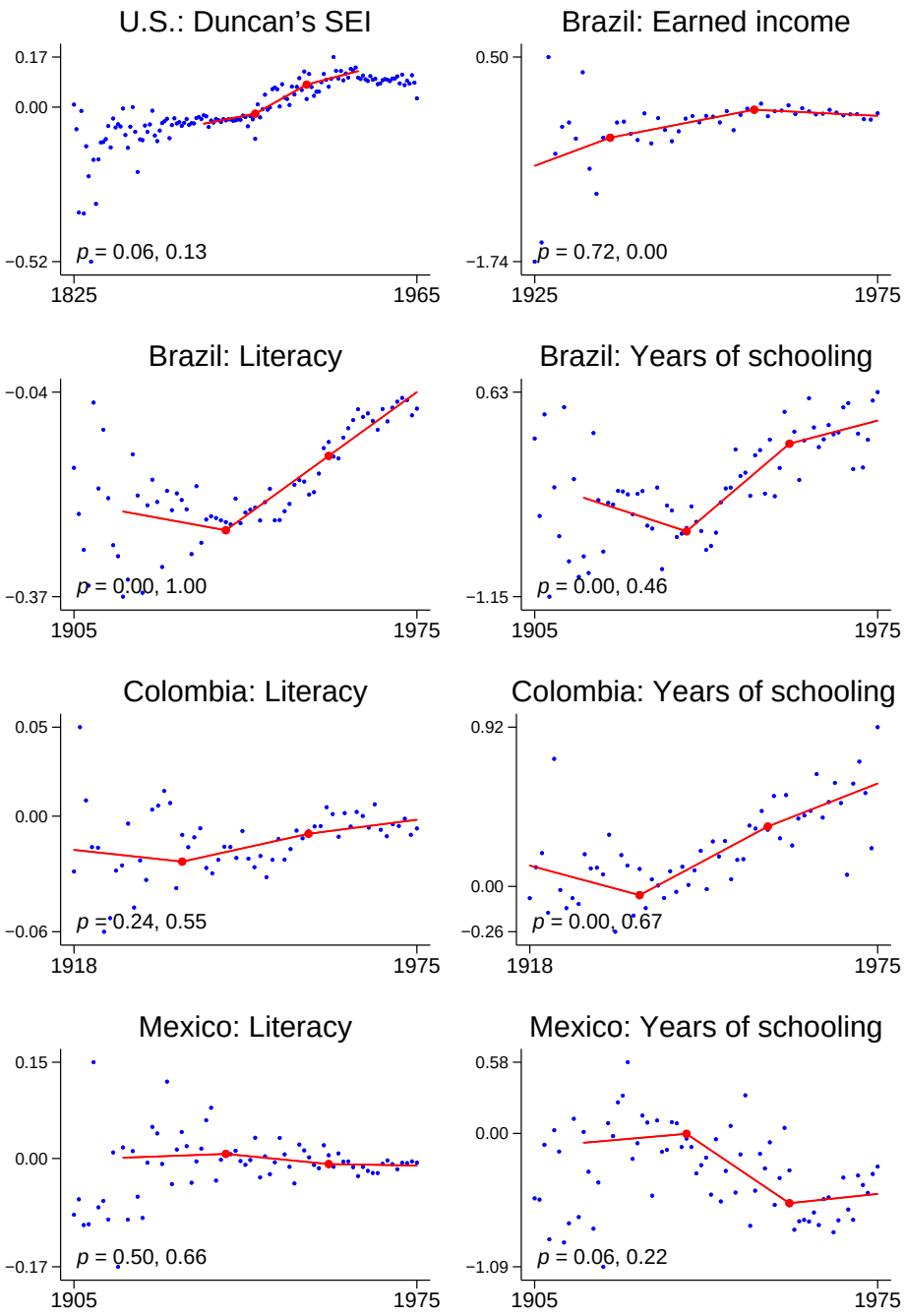
Notes: See notes for Figure 5.

Figure A4: Replication and extension of Bleakley (2010) Figure 4: Model with polynomial time controls, fit to expanded data set, alternative outcome measures, excluding women and (in U.S.) blacks



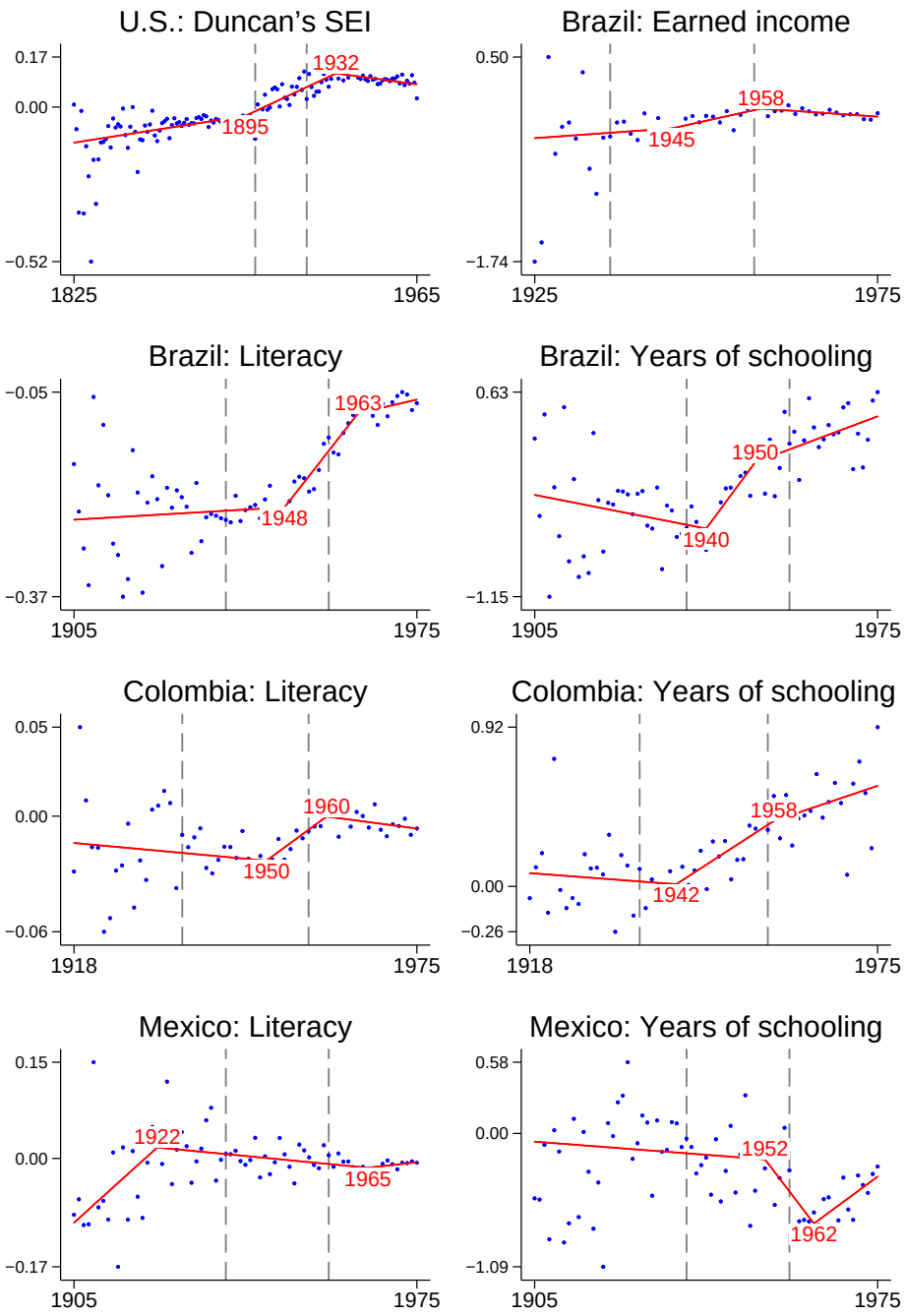
Notes: See notes for Figure 3.

Figure A5: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, fixed kink dates, alternative outcome measures, excluding women and (in U.S.) blacks



Notes: See notes for Figure 4.

Figure A6: Replication and extension of Bleakley (2010) Figure 4: Model with linear spline generalization of step function, flexible kink dates, alternative outcome measures, excluding women and (in U.S.) blacks



Notes: See notes for Figure 5.