*Malaria, Race, and Inequality: Evidence from the Early*

*1900s US South*

Online Appendix

# Additional Tables & Figures

Table A.1

Summary Statistics: Matched Sample vs. Complete Count Census

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | IPUMS Full Count | | | Linked Census Sample | | |
|  | Mean | Median | SD | Mean | Median | SD |
| Demographic Variables  Race = Black | 0.255 | 0.000 | 0.436 | 0.157 | 0.000 | 0.363 |
| Age in 1940 Census | 37.330 | 36.000 | 9.676 | 36.782 | 35.000 | 9.542 |
| Individual Level Data  Years of Education | 7.573 | 8.000 | 3.789 | 8.154 | 8.000 | 3.671 |
| Obtained Greater Than 8 Years of Schooling | 0.508 | 1.000 | 0.500 | 0.576 | 1.000 | 0.494 |
| Wage/Salary Income | 667.910 | 408.000 | 806.601 | 720.516 | 480.000 | 835.592 |
| Income, Adjusted for Self-Employed Earnings | 837.777 | 601.146 | 774.869 | 913.378 | 720.000 | 790.274 |
| Self Employed | 0.262 | 0.000 | 0.440 | 0.264 | 0.000 | 0.441 |
| *Occupations:*  Farm Owner | 0.079 | 0.000 | 0.269 | 0.084 | 0.000 | 0.277 |
| Farm Tenant | 0.108 | 0.000 | 0.310 | 0.096 | 0.000 | 0.295 |
| Farm Laborer | 0.083 | 0.000 | 0.276 | 0.085 | 0.000 | 0.279 |
| Blue Collar | 0.486 | 0.000 | 0.500 | 0.460 | 0.000 | 0.498 |
| White Collar | 0.190 | 0.000 | 0.393 | 0.230 | 0.000 | 0.421 |
| Not In Labor Force | 0.059 | 0.000 | 0.236 | 0.049 | 0.000 | 0.215 |
| *Migration:*  Migrated Across States | 0.321 | 0.000 | 0.467 | 0.292 | 0.000 | 0.454 |
| Migrated Across Counties | – | – | – | 0.619 | 1.000 | 0.486 |
| Observations |  | 9,540,300 |  |  | 1,790,568 |  |

*Notes*: Summary statistics presented for white and Black males between the ages of 23–57 (inclusive) born in the South. Full Count refers to the 1940 full count census (Ruggles et al. 2021). Linked Census Sample is the 1940 census sample linked across prior census years. Observations with missing education data are excluded.

*Source*: Authors’ calculations.

Table A.2

Impact of Malaria Eradication on Years of Education

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent Variable: Years of Education | | | | | |
|  | Pooled Across Races | | | By Race | |
|  | (1) | (2) | (3) | (4) | (5) |
| Malaria Mortality Exposure \* Treated [*β*] | 0.126∗∗ | 0.109∗∗ | 0.066 |  |  |
|  | (0.054) | (0.053) | (0.055) |  |  |
| Malaria Mortality Exposure \* Treated \* (Race = White) [*δw*] |  |  |  | 0.135∗∗ | 0.107∗ |
|  |  |  |  | (0.056) | (0.058) |
| Malaria Mortality Exposure \* Treated \* (Race = Black) [*δb*] |  |  |  | -0.001 | -0.105 |
|  |  |  |  | (0.078) | (0.081) |
| White - Black Difference [*δw* - *δb*] |  |  |  | 0.135∗ | 0.211∗∗∗ |
|  |  |  |  | (0.072) | (0.078) |
| Observations | 1,790,568 | 1,790,555 | 1,710,524 | 1,790,555 | 1,710,524 |
| Clusters | 1,398 | 1,398 | 1,326 | 1,398 | 1,326 |
| County Fixed Effects | Yes |  |  |  |  |
| Birth Year Bin Fixed Effects | Yes |  |  |  |  |
| County \* Race Fixed Effects |  | Yes | Yes | Yes | Yes |
| Birth Year Bin \* Birth State \* Race Fixed Effects |  | Yes | Yes | Yes | Yes |
| Controls |  |  | Yes |  | Yes |

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes*:Controls include 1910 male unemployment rate in childhood county interacted with birth year bin and 1890 non-malaria mortality per 1,000 population in childhood county interacted with birth year bin. Robust standard errors are clustered at the childhood county level.

*Source*: Author’s calculations. The table presents results from estimating Equations (1) and (2).

Table A.3

Impact of Malaria Eradication on Log Income (Not Adjusted)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dependent Variable: Log Wage Income | | | | | | |
|  | Pooled Across Races | | | By Race | | |
|  | (1) | (2) | (3) | | (4) | (5) |
| Malaria Mortality Exposure \* Treated [*β*] | 0.050∗∗∗ | 0.011 | 0.014 | |  |  |
|  | (0.009) | (0.012) | (0.013) | |  |  |
| Malaria Mortality Exposure \* Treated \* (Race = White) [*δw*] |  |  |  | | 0.012 | 0.018 |
|  |  |  |  | | (0.013) | (0.014) |
| Malaria Mortality Exposure \* Treated \* (Race = Black) [*δb*] |  |  |  | | 0.004 | -0.001 |
|  |  |  |  | | (0.018) | (0.021) |
| White - Black Difference [*δw* - *δb*] |  |  |  | | 0.008 | 0.019 |
|  |  |  |  | | (0.020) | (0.022) |
| Observations | 1,221,464 | 1,221,441 | 1,145,137 | | 1,221,441 | 1,145,137 |
| Clusters | 1,398 | 1,398 | 1,319 | | 1,398 | 1,319 |
| County Fixed Effects | Yes |  |  | |  |  |
| Birth Year Bin Fixed Effects | Yes |  |  | |  |  |
| County \* Race Fixed Effects |  | Yes | Yes | | Yes | Yes |
| Birth Year Bin \* Birth State \* Race Fixed Effects |  | Yes | Yes | | Yes | Yes |
| Controls |  |  | Yes | |  | Yes |

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes*:Controls include 1910 male unemployment rate in childhood county interacted with birth year bin, 1890 non-malaria mortality per 1,000 population in childhood county interacted with birth year bin, and Rosenwald school exposure in childhood county. Robust standard errors are clustered at the childhood county level.

*Source*: Author’s calculations. The table presents results from estimating Equations (1) and (2).

Table A.4

Impact of Malaria Eradication on Schooling and Income: State of Birth Level Variation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Dependent Variable: | | | |
|  | Middle School Completion | | Log Income | |
|  | (1) | (2) | (3) | (4) |
| Malaria Mortality Exposure (State) \* Treated \* (Race = White) [*δw*] | 0.031 | 0.030 | 0.114∗∗∗ | 0.110∗∗∗ |
|  | (0.023) | (0.022) | (0.032) | (0.032) |
| Malaria Mortality Exposure (State) \* Treated \* (Race = Black) [*δb*] | -0.087∗ | -0.060 | 0.165∗∗∗ | 0.181∗∗∗ |
|  | (0.046) | (0.043) | (0.035) | (0.038) |
| White - Black Difference [*δw* - *δb*] | 0.118∗∗ | 0.090∗ | -0.051 | -0.071∗∗ |
|  | (0.047) | (0.048) | (0.036) | (0.032) |
| Observations | 5,819,040 | 5,737,966 | 5,322,227 | 5,242,608 |
| Clusters | 49 | 49 | 49 | 49 |
| State of Birth \* Race Fixed Effects | Yes | Yes | Yes | Yes |
| Birth Year Bin \* Race Fixed Effects | Yes | Yes | Yes | Yes |
| Controls |  | Yes |  | Yes |

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes*:The sample includes Northern-born individuals. In all columns, malaria mortality exposure is redefined to vary at the state of birth level. All columns include controls for 1910 male unemployment rate in childhood county interacted with birth year bin and 1890 non-malaria mortality per 1,000 population in childhood county interacted with birth year bin. Columns (3)-(4) additionally include controls for Rosenwald exposure in childhood county. Robust standard errors are clustered at the childhood county level.

*Source*: Authors’ calculations. The table presents results from estimating Equation (2).

Table A.5

Robustness to Alternative Specifications: Impact of Malaria Eradication on Income

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent Variable: Log of Income, Adjusted for Self-Employed Earnings | | | | | |
|  | Baseline | Unbinned Treatment | Non-Parametric Mortality | Malaria Ecology | Alt. Dep. Var |
|  | (1) | (2) | (3) | (4) | (5) |
| Panel A: Without Controls |  |  |  |  |  |
| Malaria Mortality Exposure \* Treated \* (Race = White) [*δw*] | 0.013 | 0.012 | -0.003 | -0.006 | 0.037∗∗∗ |
|  | (0.009) | (0.009) | (0.009) | (0.050) | (0.012) |
| Malaria Mortality Exposure \* Treated \* (Race = Black) [*δb*] | 0.005 | 0.003 | -0.021 | -0.052 | 0.000 |
|  | (0.013) | (0.013) | (0.015) | (0.082) | (0.022) |
| White - Black Difference [*δw* - *δb*] | 0.008 | 0.009 | 0.018 | 0.046 | 0.037 |
|  | (0.014) | (0.014) | (0.016) | (0.088) | (0.023) |
| Observations | 1,628,176 | 1,628,176 | 1,628,176 | 1,574,918 | 513,702 |
| Clusters | 1,398 | 1,398 | 1,398 | 1,340 | 1,392 |
| Panel B: With Controls |  |  |  |  |  |
| Malaria Mortality Exposure \* Treated \* (Race = White) [*δw*] | 0.020∗∗ | 0.018∗∗ | -0.000 | -0.002 | 0.042∗∗∗ |
|  | (0.010) | (0.009) | (0.009) | (0.054) | (0.015) |
| Malaria Mortality Exposure \* Treated \* (Race = Black) [*δb*] | 0.004 | 0.005 | -0.026∗ | -0.075 | 0.019 |
|  | (0.015) | (0.014) | (0.015) | (0.086) | (0.024) |
| White - Black Difference [*δw* - *δb*] | 0.016 | 0.014 | 0.025 | 0.072 | 0.023 |
|  | (0.015) | (0.015) | (0.016) | (0.089) | (0.023) |
| Observations | 1,532,940 | 1,532,940 | 1,532,940 | 1,528,058 | 478,034 |
| Clusters | 1,319 | 1,319 | 1,319 | 1,313 | 1,316 |
| County \* Race FE | Yes | Yes | Yes | Yes | Yes |
| Birth Year Bin \* Birth State \* Race FE | Yes | Yes | Yes | Yes | Yes |

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes*: Panel (A) includes results without controls, and panel (B) includes controls. Controls include 1910 male unemployment rate in childhood county interacted with birth year bin, 1890 non-malaria mortality per 1,000 population in childhood county interacted with birth year bin, and Rosenwald school exposure in childhood county. Column (2) redefines *treatt*: the variable is now given by a modified version of (4), with *t* replaced by year of birth. Column (3) redefines mortality as a binary variable that equals one if a county has above median malaria mortality. Column (4) replaces malaria mortality with the Malaria Ecology Index. Column (5) adjusts income for cost of living. Robust standard errors are clustered at the childhood county level.

*Source*: The table presents results from estimating Equation (2). Column (1) replicates the baseline regression results displayed in Table 3.

Table A.6

Robustness to Census Matching: Impact of Malaria Eradication on Income

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent Variable: Log of Income, Adjusted for Self-Employed Earnings  2-year | | | | | |
|  | Baseline | 5-year NYSIIS | 2-year Non-  NYSIIS | 0-year  NYSIIS | Weighted |
|  | (1) | (2) | (3) | (4) | (5) |
| Panel A: Without Controls |  |  |  |  |  |
| Malaria Mortality Exposure \* Treated \* (Race = White) [*δw*] | 0.013 | 0.006 | 0.013 | 0.013 | 0.007 |
|  | (0.009) | (0.010) | (0.010) | (0.009) | (0.009) |
| Malaria Mortality Exposure \* Treated \* (Race = Black) [*δb*] | 0.005 | 0.005 | 0.004 | 0.006 | 0.004 |
|  | (0.013) | (0.016) | (0.015) | (0.010) | (0.013) |
| White - Black Difference [*δw* - *δb*] | 0.008 | 0.001 | 0.009 | 0.007 | 0.003 |
|  | (0.014) | (0.017) | (0.015) | (0.011) | (0.014) |
| Observations | 1,628,176 | 1,198,900 | 1,468,832 | 2,480,676 | 1,628,176 |
| Clusters | 1,398 | 1,398 | 1,398 | 1,398 | 1,398 |
| Panel B: With Controls |  |  |  |  |  |
| Malaria Mortality Exposure \* Treated \* (Race = White) [*δw*] | 0.020∗∗ | 0.014 | 0.020∗ | 0.016 | 0.013 |
|  | (0.010) | (0.011) | (0.011) | (0.010) | (0.010) |
| Malaria Mortality Exposure \* Treated \* (Race = Black) [*δb*] | 0.004 | 0.012 | 0.005 | 0.000 | 0.003 |
|  | (0.015) | (0.018) | (0.016) | (0.012) | (0.015) |
| White - Black Difference [*δw* - *δb*] | 0.016 | 0.002 | 0.015 | 0.016 | 0.010 |
|  | (0.015) | (0.018) | (0.016) | (0.012) | (0.016) |
| Observations | 1,532,940 | 1,124,863 | 1,380,295 | 2,344,912 | 1,532,940 |
| Clusters | 1,319 | 1,319 | 1,319 | 1,319 | 1,319 |
| County \* Race FE | Yes | Yes | Yes | Yes | Yes |
| Birth Year Bin \* Birth State \* Race FE | Yes | Yes | Yes | Yes | Yes |

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes*: Panel (A) includes results without controls, and panel (B) includes controls. Controls include 1910 male unemployment rate in childhood county interacted with birth year bin, 1890 non-malaria mortality per 1,000 population in childhood county interacted with birth year bin, and Rosenwald school exposure in childhood county. Column (2) requires observations to be unique within 5 years in its own dataset. Column (3) uses exact names rather than standardized names. Column (4) only requires an observation to be unique in its own year of birth in its dataset. Column (5) uses inverse probability weights for the demographic characteristics used to match (year of birth, race, and state of birth) individuals across census years. Robust standard errors are clustered at the childhood county level. Robust standard errors are clustered at the childhood county level.

*Source*: Authors’ calculations. The table presents results from estimating Equation (2). Column (1) replicates the baseline regression results displayed in Table 3.

Diagram

Description automatically generated

## Figure A.1

## Malarious Areas of the United States

*Notes*: This figure shows the distribution of malaria in the United States.

*Source*: “Elimination of Malaria in the United States (1947 – 1951),” *Centers for Disease Control and Prevention*, [https://www.cdc.gov/malaria/ about/history/elimination](https://www.cdc.gov/malaria/about/history/elimination_us.html) [us.html.](https://www.cdc.gov/malaria/about/history/elimination_us.html)

Chart

Description automatically generated

## Figure A.2

Log Income, Adjusted for Self-Employed Earnings, and Malaria Mortality – Raw Data

*Notes*: This figure plots log income adjusted for self-employed earnings by race, birth year, and above and below median malaria mortality. Above (below) median includes all counties that reported greater (less) than median malaria mortality in 1890.

*Source*: Authors’ calculations.

# 

# B Comparison to Bleakley (2010)

In this section, we show how our sample and specification corresponds to a well-known result in the literature, Bleakley (2010). Bleakley studies the impact of malaria eradication programs on white males and finds that eradication programs are associated with rising white male productivity. Bleakley’s sample consists of white males between the ages of 35 and 55 for census years 1880-2000. As Bleakley uses IPUMS and wage data is not available until 1940, his proxies for labor productivity are occupational income score and the Duncan socioeconomic index (SEI).

Bleakley’s baseline results stem from a pre/post comparison. Bleakley compares males who were already adults when eradication campaigns began to males who were born after eradication campaigns began (i.e., “partially exposed” cohorts are not included in this section of his analysis). His baseline estimating equation is:

where *Y* is a proxy for labor productivity, is the malaria incidence in state of birth , is a vector of controls, and *α* is a constant.

We replicate Equation (B.1) on Bleakley’s publicly provided sample of data for Duncan SEI, as seen in Figure B.1. Bleakley’s baseline result shows that white males exposed to eradication campaigns have a statistically significant higher Duncan SEI.

We then adjust Bleakley’s sample and methodology over a sequence of steps to arrive at our preferred sample and specification. We first adjust the definition of “malaria incidence.” Bleakley defines malaria incidence as in an individual’s state of birth. Our preferred measure of malaria exposure is malaria mortality (i.e., malaria deaths per 1,000 population). Making this adjustment to malaria incidence definition while utilizing Bleakley’s sample and specification makes the results noisier but nearly identical in magnitude, as seen in the second coefficient in Figure B.1.

Our methodology uses the 1940 full count census. We match individuals in the full count census to their childhood census, thus obtaining the linked sample discussed in the “Linked Census Data” section. The remaining comparisons will make adjustments to the 1940 Census in order to transition from Bleakley’s methodology of repeated cross sections to our methodology of linked Census records. Our sample across all remaining specifications will include white males aged 23 to 57 and our specification will follow our baseline specification in Equation (1):

where *Y* is a proxy for labor productivity, is malaria mortality in geography (either state or county of birth), is length of time exposed to eradication campaigns, is a vector of controls, is a birth cohort fixed effect, and is a birth geography (state/county) fixed effect.

Before turning to our matched sample, we first replicate Bleakley’s results on the 1940 full count Census in the third coefficient of Figure B.1. Malaria exposure () is state of birth malaria mortality and we include birth cohort fixed effects and state of birth fixed effects. This specification is roughly analogous to Bleakley’s baseline specification. We find positive and statistically significant results that are slightly smaller in magnitude than Bleakley’s baseline, but are more precisely estimated and fall within his initial confidence interval. The fourth coefficient in Figure B.1 turns to our linked sample. We again find a result that is similar to Bleakley’s baseline result.

Our preferred sample focuses on males born in the south since this was the region in the United States where malaria was problematic and where eradication efforts were focused in the 1920s. In addition, since we have linked census records, we are able to use finer levels of malaria mortality exposure than Bleakley was able to employ. In particular, we can define malaria mortality in the county of birth. We use this county-level variation for Southern born males in the fifth coefficient of Figure B.1. We find nearly identical results to Bleakley’s baseline when using this finer variation but maintaining state of birth fixed effects.

In the sixth coefficient of Figure B.1, we include county of birth fixed effects in place of state of birth fixed effects. While we do see a positive and statistically significant coefficient, it is much closer to zero than initially found in Bleakley’s result. This suggests that much of the positive coefficient found across coefficients 1-5 can be explained by across state convergence.

In coefficients 7 and 8, we are able to include even stronger fixed effects. In particular, we adjust the standard birth cohort fixed effects that were implemented in coefficients 1-6 to be birth cohort by state of birth fixed effects. In coefficient 8, we add our set of controls (1890 non-malaria mortality, 1910 male unemployment, and Rosenwald exposure), representing our preferred specification. Allowing for different trends by state of birth and birth cohort can explain the remaining positive effect. The remaining two coefficients are even closer to zero with tight standard errors. Thus while we are initially able to find similar results to Bleakley, this result does not hold up to the finer levels of variation and fixed effects that we are able to include.

We perform the same exercise using occupational score as the proxy for labor productivity. The results are displayed in Figure B.2. We reach a similar conclusion as in our Duncan SEI analysis. One exception is that coefficients 3 and 4 (1940 full count census and 1940 linked census) do not closely match Bleakley’s results. However, we do find results that are more aligned to his result once we include our finer county variation in coefficient 5. While we are unable to explain this deviation, we think the Duncan SEI results are a better proxy for labor productivity since Duncan SEI includes information about education that is not included in occupational score.

Chart, box and whisker chart

Description automatically generated

Figure B.1

Duncan SEI Coefficient Comparisons

*Notes*: This figure plots coefficients from various regressions with Duncan SEI as the dependent variable. The sample in all regressions is white males. Coefficient 1 is Bleakley’s baseline result using repeated cross sections of Census data. Coefficient 2 uses Bleakley’s data but changes the definition of malaria incidence. Coefficient 3 uses the 1940 full count census with state of birth malaria mortality, and state of birth and birth cohort fixed effects. Coefficient 4 uses the 1940 matched census sample with state of birth malaria mortality, and state of birth and birth cohort fixed effects. Coefficient 5 uses the 1940 matched census sample for Southern states only with county of birth malaria mortality, and state of birth and birth cohort fixed effects. Coefficient 6 uses the 1940 matched census sample for Southern states only with county of birth malaria mortality, and county of birth and birth cohort fixed effects. Coefficient 7 uses the 1940 matched census sample for Southern states only with county of birth malaria mortality, and county of birth and birth cohort by state of birth fixed effects. Coefficient 8 adds controls for 1890 non-malaria mortality, 1910 male unemployment, and Rosenwald school exposure.

*Source*: Author’s calculations; Bleakley (2010).

Chart, box and whisker chart

Description automatically generated

Figure B.2

Occupational Score Coefficient Comparisons

*Notes*: This figure plots coefficients from various regressions with Occupational Score as the dependent variable. The sample in all regressions is white males. Coefficient 1 is Bleakley’s baseline result using repeated cross sections of Census data. Coefficient 2 uses Bleakley’s data but changes the definition of malaria incidence. Coefficient 3 uses the 1940 full count census with state of birth malaria mortality, and state of birth and birth cohort fixed effects. Coefficient 4 uses the 1940 matched census sample with state of birth malaria mortality, and state of birth and birth cohort fixed effects. Coefficient 5 uses the 1940 matched census sample for Southern states only with county of birth malaria mortality, and state of birth and birth cohort fixed effects. Coefficient 6 uses the 1940 matched census sample for Southern states only with county of birth malaria mortality, and county of birth and birth cohort fixed effects. Coefficient 7 uses the 1940 matched census sample for Southern states only with county of birth malaria mortality, and county of birth and birth cohort by state of birth fixed effects. Coefficient 8 adds controls for 1890 non-malaria mortality, 1910 male unemployment, and Rosenwald school exposure.

*Source*: Author’s calculations; Bleakley (2010).

# Variable Definitions

## Details on Income Adjustments

We follow Collins and Wanamaker (2014) and adjust incomes as reported in the 1940 complete count census to include income from self-employment. The 1960 5 percent census contains data on income from both wages and self-employment. The adjustment procedure is as follows: First, we compute the ratio of average self-employed income to average wage-earners income in 1960 at an occupation x region x race level, as shown below:

where indexes individual, indexes occupation, indexes census region, and indexes race. is the calculated adjustment factor. denotes the total number of individuals in a particular occupation x region x race cell. and denote income from self-employment and wages, respectively.

Second, we compute average wage-earners income in 1940:

where the common variables and indices are exactly as above. denotes the average wage income in 1940 at an occupation x region x race level.

Finally, we impute self-employed income in 1940 as the product of (C.1) and (C.2):

## Rosenwald Exposure Calculation

The Rosenwald exposure variable estimates the average Rosenwald school coverage experienced by a student aged 7-13. We replicate Aaronson and Mazumder (2011) and calculate exposure as stated below:

where indexes year of birth, indexes county of residence, and indexes year. is the exposure measure for a student born in year living in county . is the number of Rosenwald teachers in a particular county in year . This data was obtained from digitized Rosenwald school records, made available by Aaronson & Mazumder. is the number of rural Blacks aged 7-17, sourced from census records and interpolated for non-census years. Finally, we assume a class size of 45 students, which was standard for the time. , with a value of zero indicating no Rosenwald presence in a county for a particular cohort, and one indicating full coverage i.e., every eligible Black child had access to a Rosenwald school.

## Neighbor-Based Segregation Measure

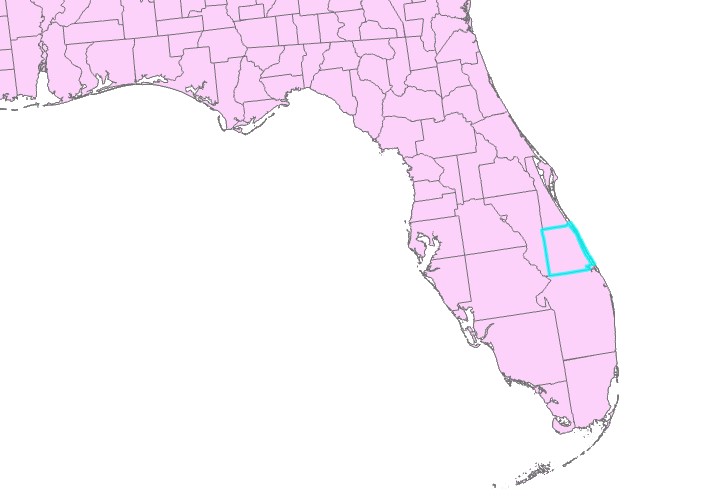
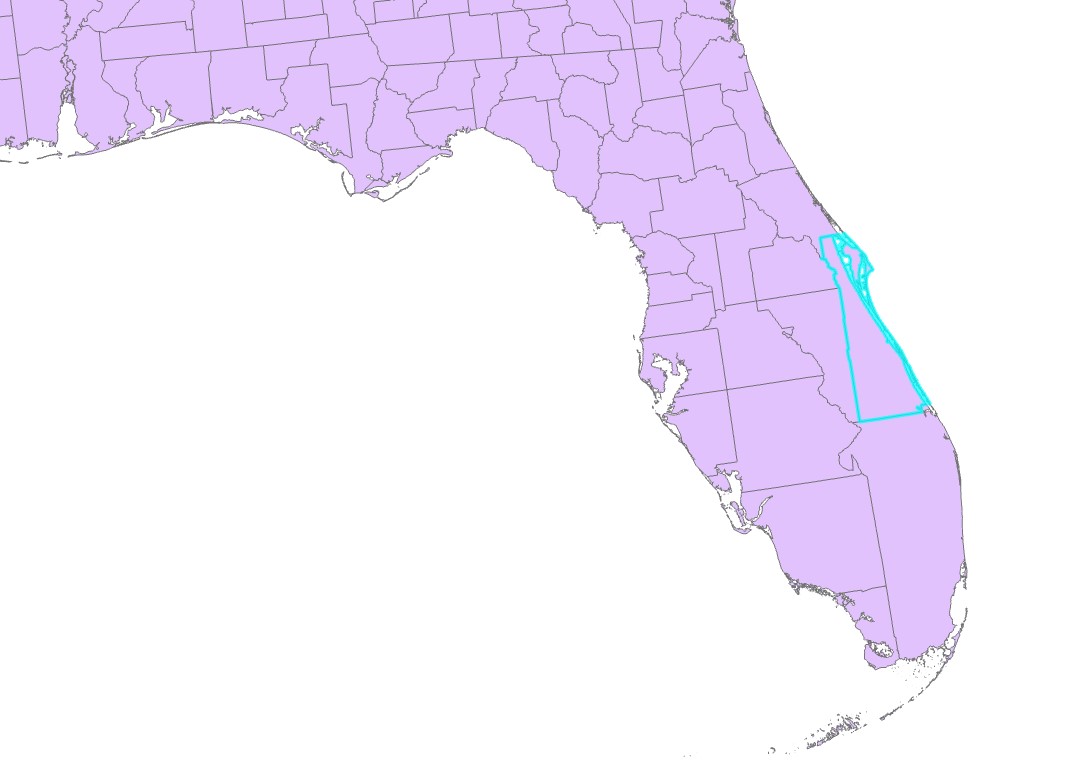
Data on the neighbor-based segregation index as calculated in Logan and Parman (2017) has been made publicly available by the authors.[[1]](#footnote-1) For a particular census year, the index compares the number of Black households living next to opposite race neighbors relative to the expected number under complete segregation and no segregation (i.e., random assignment). It is calculated as follows:

where indexes county and denotes Blacks. is the county-level value of the segregation index. refers to the observed number of Black households in county *c* with an opposite race next door neighbor. is the expected number for under complete *integration* (i.e., if neighbors were randomly assigned across races). is the expectation value of under complete *segregation* (i.e., if Blacks only lived next to Blacks). As is clear from (C.4), is increasing in the level of segregation. When , equals zero. Conversely, when , equals one.[[2]](#footnote-2)

## County Boundary Adjustments

We construct county crosswalks for the 1880-1940 decadal census years. County boundary data is sourced from the US Census Bureau’s TIGER/Line shape files, available through the IPUMS National Historical Geographic Information System (NHGIS) website.[[3]](#footnote-3) We import the shapefiles into GIS software and intersect county boundaries across census years to obtain the fraction of a county’s land area in any given (i.e., linking) year that intersects with a county from a previous or future (i.e., master) census year. The approach allows us to assign a county-level variable sourced from any particular census year (the master counties) to counties in all future and previous census years (the assignment counties) based on how assignment county boundaries overlay on top of master county boundaries.

An example illustrates our approach. Figure C.1 below shows the county boundaries for Florida from the 1890 and 1910 census years, respectively. The county outlined in blue in panel (b) is St. Lucie County. St. Lucie was created in 1905 from parts of Brevard County, which is outlined in blue in panel (a).



(a) 1890 (b) 1910

Figure C.1

Florida County Boundaries, 1890 and 1910 Census Years

Since we source malaria data from the 1890 vital statistics, we can only compute malaria mortality according to the 1890 county boundary definitions. Thus, we cannot use this data to determine malaria exposure for an individual who reports being born in St. Lucie county in the 1910 census. However, from the crosswalk, we obtain that 100 percent of St. Lucie county’s land area came from historical Brevard county. We then assign pre-campaign malaria mortality for St. Lucie as 100 percent of mortality recorded in Brevard in 1890. In this particular case, since Brevard had a malaria mortality rate of zero in 1890, St. Lucie is also assigned a mortality value of zero. We follow the exact same approach for every other county across all census years.

1. We are grateful to the authors for making their dataset publicly available. [↑](#footnote-ref-1)
2. For more information, see Logan and Parman (2017). [↑](#footnote-ref-2)
3. Available at: [https://www.nhgis.org/.](https://www.nhgis.org/) [↑](#footnote-ref-3)