

# *Gender, race, and heterogeneous effects of epidemic malaria on human capital and income*

Article

Published Version

Rawlings, S. B. (2016) Gender, race, and heterogeneous effects of epidemic malaria on human capital and income. *Economic Development and Cultural Change*, 64 (3). pp. 509-543. ISSN 1539-2988 doi: <https://doi.org/10.1086/684965>  
Available at <http://centaur.reading.ac.uk/47246/>

It is advisable to refer to the publisher's version if you intend to cite from the work.

To link to this article DOI: <http://dx.doi.org/10.1086/684965>

Publisher: The University of Chicago Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Gender, Race, and Heterogeneous Effects of Epidemic Malaria on Human Capital and Income

SAMANTHA B. RAWLINGS

University of Reading

## I. Introduction

Malaria is one of the most pervasive diseases in developing countries today. Despite a large decline in incidence of the disease since 2000, an estimated 207 million people contracted it in 2012 with around 627,000 deaths, most of these among African children under age 5 (WHO 2013). The overwhelming majority of malaria cases occur in Africa, with sub-Saharan Africa containing both a climate suitable for mosquitoes and native mosquito species that are particularly suited to transmitting the disease (Gallup and Sachs 2001). Although there has been huge progress in combating the disease, there are concerns that climate change–induced rises in average temperatures may not only increase the malaria burden in areas where there is currently transmission but also spread the disease to currently nonendemic areas such as the East African Highlands (e.g., Lindsay and Martens 1998; Martens et al. 1999; van Lieshout et al. 2004; Siraj et al. 2014). It has been argued that malaria contributes to lower income growth through a number of channels, such as depressing adult work productivity, increasing premature mortality, and affecting saving and investment (Sachs and Malaney 2002). However, the existing cross-country literature on malaria and income growth debates whether it is malaria itself that matters for human capital attainment or whether malaria is correlated with other causes of underdevelopment (e.g., Sachs and Malaney 2002; Rodrik, Subramanian, and Trebbi 2004). One argument is that malaria does not affect income directly but only through its impact on other factors such as political institutions (e.g., Acemoglu, Johnson, and Robinson 2003). The empirical issue is then one of finding an exogenous source of variation in malaria from which to investigate the link between it and income growth.

In recent years, a rash of microeconomic studies of the impact of early life exposure to malaria on adult health, human capital attainment, and income has emerged, with two distinct empirical approaches. The first relies on instru-

I thank Sonia Bhalotra, Jon Temple, Marcos Vera Hernandez, and participants at the Centre for Study of African Economies Conference for useful comments and suggestions. Contact the author at [s.b.rawlings@reading.ac.uk](mailto:s.b.rawlings@reading.ac.uk).

Electronically published January 29, 2016  
© 2016 by The University of Chicago. All rights reserved. 0013-0079/2016/6403-0002\$10.00

mental variable strategies to isolate the causal effect of malaria using historical US data (Hong 2007; Barreca 2010), finding negative effects on both health and human capital. The second approach investigates the extent to which human capital attainment was affected by malaria eradication campaigns in the Americas (Bleakley 2010), Sri Lanka and Paraguay (Lucas 2010), India (Cutler et al. 2010), Mexico (Venkataramani 2010), and Taiwan (Chang et al. 2011). To identify effects, these papers exploit variation in preradication infection rates combined with differential exposure to the campaigns across cohorts. They tend to find eradicating malaria increases income, consumption, and IQ scores, with little to no effect on schooling (e.g., Bleakley 2010), with one study finding that the negative consequences of epidemic malaria are larger than the positive contribution to human capital conferred by malaria eradication (Lucas 2010). Larger effects are found for males than females, suggesting that they benefit more from malaria eradication. This is consistent with a biomedical literature suggesting that young males have less resistance to environmental shocks than females. A recent paper by Kuecken, Thuilliez, and Valfort (2015) combines the two approaches to investigate the impact of reductions in malaria on short-term education outcomes of children, using the introduction of the Roll Back Malaria campaign alongside the use of geographic, climatic, and genetic variables to instrument for precampaign malaria; it finds that reductions in malaria have a positive effect on the years of schooling completed by primary schoolchildren.

Aside from Kuecken et al. (2015), these papers focus on areas of the world in which the dominant strain of malaria is *Plasmodium vivax*, a less virulent and less fatal strain of malaria than *Plasmodium falciparum* (Mendis et al. 2001), which is the leading cause of malaria deaths worldwide (WHO 2008). Perhaps because of this, there is little discussion in this particular literature of the role of mortality selection, with a focus on the reduction in scarring that can be attributed to malaria eradication. However, heterogeneous scarring and mortality selection effects have been found elsewhere, such as in the relationship between early disease environment and offspring health (Almond, Currie, and Herrmann 2012) and the long-term effects of being a Civil War prisoner of war in the United States (Costa 2012).

A complementary literature looks at the long-term impacts of exposure to a variety of exogenous environmental stresses, such as flu epidemics (Almond 2006; Kelly 2011), famine (Chen and Zhou 2007; Meng and Qian 2009; Almond et al. 2010), rainfall shocks (Maccini and Yang 2009), civil war (Domingues 2010; Galdo 2010; León 2010), or other one-off events such as the Chernobyl disaster (Almond, Edlund, and Palme 2009). Negative effects on literacy, income, schooling, labor supply, and height have been found in

this literature, although results are not always consistent or statistically significant. More attention has been paid to the in utero rather than the postnatal period (Almond and Currie 2011), motivated in part by the so-called fetal origins hypothesis (Barker 1992, 1997; Barker and Levy 1994). The idea is that exposure to a disaster when in the womb affects child development and that this permanently inhibits attainment of human capital. Again, there is little discussion of the impact of mortality selection in this line of literature (Valente 2015).

This article investigates the impact of an exogenous, isolated epidemic of malaria occurring in two states of Brazil (Ceará and Rio Grande do Norte) from 1938 to 1940. It exploits both regional and cohort variation in exposure to the epidemic, using as a control group individuals classified as unexposed to the epidemic. State and year fixed effects as well as state-specific trends are included to account for any potential inherent differences between the treatment and control groups. I am unable to disentangle the effect of disease as distinct from epidemic-associated trauma such as loss of income and undernutrition; results are thus interpreted as the reduced-form effect of exposure to a large, negative, environmental shock.

Results suggest heterogeneous effects according to gender and race. Estimates vary according to year and state of exposure, but in general I find that, on average, for nonwhite women exposure to the epidemic resulted in a cohort of individuals with higher levels of schooling and income, while for white women estimated impacts of the epidemic on schooling and income were negative. Evidence for men is weaker but suggests that, on average, exposure to the epidemic results in a cohort of individuals with higher levels of schooling and income, regardless of race. These results are consistent with a model of competing mortality selection and scarring effects, as discussed by Bozzoli, Deaton, and Quintana-Domeque (2009), with apparent heterogeneity in which effect dominates according to gender and race.

Gender differences may stem from the fact that females have greater resistance to infectious disease (Waldron 1983). Young males are more likely to die and therefore as a group have relatively higher endowments of human capital compared to their nonexposed peers. Women, however, who are more resistant to environmental shocks and therefore as a group have lower mortality, may be selected or scarred, depending on their socioeconomic background. In this context, race serves as a proxy for socioeconomic status, with higher mortality rates expected for individuals with lower socioeconomic status (Worrall, Basu, and Hanson 2005; Dickinson et al. 2012).

As an example, directly comparing coefficients for nonwhite and white women for Ceará, I find that exposure to the epidemic led to a 2.0 percentage

point increase in the probability of having primary schooling for nonwhite women. In contrast, the corresponding estimated effect for white women is a 2.7 percentage point decrease in the likelihood of having primary schooling. These effects are sizable, ranging from 12.3% to 39.4% of the gender- and race-specific sample means of these variables.

This article contributes to the existing literature in the following ways: It focuses on a transitory negative shock rather than a permanent improvement in the disease environment as previously investigated. It examines whether the effect is heterogeneous across groups and discusses selection versus scarring effects when estimating long-term impacts of environmental shocks. The rest of the article is structured as follows. Section II describes key features of malaria as a disease and outlines the epidemic in Brazil. Section III discusses the data used and provides some descriptive analysis, Section IV outlines the methodology employed, Section V presents results, and Section VI presents robustness checks. Finally, Section VII concludes.

## II. Malaria

### *Causes and Consequences of Malaria*

Malaria is caused by the infection of an individual's red blood cells by a malaria parasite and is transmitted by female *Anopheles* mosquitoes, known as malaria vectors. To transmit malaria, the vector first ingests the blood of an infected human. The vector is not infectious immediately, since the parasite must develop into a sporozoite before it can infect the human, and becomes infectious around 10–21 days after ingestion of the parasite, passing on the disease to its next human victim upon feeding. Thus, killing the vector before this occurs is an effective way of halting spread of the disease.

Successful transmission of malaria depends on environmental factors such as precipitation and temperature, as well as the species of mosquito. Precipitation affects malaria transmission since bodies of water are needed for mosquito breeding to take place. Temperature determines malaria transmission since at temperatures below 16°–18°C, sporozoite development is limited. The effectiveness of a malaria vector in transmitting malaria is known as its *efficiency*; this is determined by its longevity and preference for feeding on humans (Bremen 2001). The most efficient species of mosquito in transmitting malaria belong to the *Anopheles Gambiae* complex, which has the highest rates of sporozoite development and a long life span and favors feeding on humans (Bremen 2001). It is found exclusively in sub-Saharan Africa, where its presence contributes to the prevailing high malaria infection rates (Gallup and Sachs 2001).

Different malaria parasites have differing levels of virulence; the most serious and lethal parasite is that of *Plasmodium falciparum*, the burden of which

is largely borne in Africa, where around 70% of cases of this type of infection occur (Snow et al. 2005). Weaker strains such as *Plasmodium vivax* are more common outside of Africa, where they account for more than 50% of cases (Mendis et al. 2001); although less lethal, these weaker strains still negatively affect an individual's health. Thus, symptoms of malaria vary in intensity according to the parasite. Such symptoms include high fever, headache, severe chills or rigor, swelling of the brain, seizure, as well as vomiting, cough, or diarrhoea (WHO 2000; Holding and Snow 2001). Pregnant women are more at risk of contracting malaria than nonpregnant women, although the mechanisms through which this occurs are not well understood (Lindsay et al. 2000). Malaria during pregnancy can lead to transplacental infection, which occurs when malaria is directly transmitted to the fetus. This is rare but more likely to occur among populations with low immunity to the disease (Brabin 1991). Malaria that is not directly transmitted to the infant still has severe consequences; malaria in pregnancy is associated with increased risk of low birth weight, fetal growth retardation, still birth, maternal anemia, and postpartum hemorrhage (Brabin 1991; Lindsay et al. 2000; WHO 2000; Holding and Snow 2001). Malaria infection at a young age may affect cognitive development and later life human capital development, either directly as a result of brain insults resulting from contracting malaria itself or through associated illness and undernutrition (Holding and Snow 2001). It has been estimated that cerebral malaria among children leads to sustained impaired cognition in 24% of cases (Dulac 2010). Malaria infection has also been shown to be associated with impaired physical growth of young children (Holding and Kitsao-Wekulo 2004). However, many of the existing studies on these effects of malaria are based on observational studies, and there is less causal evidence available.

#### *The Malaria Epidemic in Northeast Brazil, 1938–40*

The Brazilian epidemic of 1938–40 was the result of the accidental transportation of a particularly efficient foreign vector from Africa, *Anopheles Arabiensis*, into the states of Ceará and Rio Grande do Norte in the northeast region of Brazil (Soper and Wilson 1943; Parmakelis et al. 2008).<sup>1</sup> This particular species is part of the *Anopheles Gambiae* complex described above, which contributes to the high malaria burden that Africa suffers today (Morlais et al. 2005). The states invaded by *Anopheles Arabiensis* provided a habitat highly conducive to

<sup>1</sup> Note that it is not known exactly when the vector invaded Brazil or where it originated from. However, given that the vector was first found in a port town, that it is a species native to Africa, and that there was a high volume of shipping traffic between Brazil and Senegal in the 1930s, it has been hypothesized that this region is the most likely origin (Parmakelis et al. 2008; Oliveira-Ferreira et al. 2010).

its survival, breeding, and potential to transmit malaria. Temperatures in these states were never low enough to interfere with the development of the parasite, and there were many suitable breeding locations near households (Soper and Wilson 1943). Preexisting low levels of endemicity meant that the population had little immunity against the disease, which exacerbated the severity of the epidemic. Furthermore, both *Plasmodium falciparum* and *Plasmodium vivax* parasites were responsible for the epidemic (Causey, Penido, and Deane 1943), and the presence of *Plasmodium falciparum* contributed to the high fatality rate in the region (Trape and Rodgier 1996).

This particular epidemic was “unpredictable” (WHO 1998) and is treated as an exogenous health shock. The presence of the invader was discovered by public health officials, and its eradication was seen as an effective way of returning to the low malaria rates that had existed in the region. Through a joint intervention by the Brazilian government and the New York Rockefeller Foundation, *Anopheles Arabiensis* was successfully eradicated from the region. By the end of 1940, the epidemic was over, and malaria cases had returned to their preexisting levels. No *Anopheles Arabiensis* has been found in the area since (Killeen et al. 2002).

More specifically, the timeline of the invasion of Brazil by *Anopheles Arabiensis* was as follows:

- 1930–31. In March 1930 the *Anopheles Arabiensis* mosquito was first found in the shipping port of Natal, in Rio Grande do Norte in Brazil. By the end of June a minor epidemic had occurred that was limited to a small area and lasted only until the dry season. A second epidemic occurred in the early months of 1931 with around 10,000 cases; the area affected was limited, and the epidemic passed reasonably quickly.
- 1932–37. Known as the “silent period.” Unusually dry weather in 1930 and 1931 culminated in a drought in 1932, which significantly reduced the number of malaria vectors. With the return of “normal” weather conditions, the number of vectors increased and spread throughout the region, likely via transport (e.g., cars, boats). No major malaria incidents occurred because of unsuitable topographical conditions (Soper and Wilson 1943).<sup>2</sup>

<sup>2</sup> Since my identification strategy relies on a common trends assumption, it is important to consider whether rainfall may be an omitted variable and therefore a threat to my identification. However, the major cause of the timing of the epidemic was the (exogenous) transportation of mosquitoes to



1938. *Anopheles Arabiensis* settled in areas of the states containing a suitable breeding and survival environment, with many potential breeding grounds in the form of shallow pools of water. This triggered an explosive epidemic of malaria in the Jaguaribe River Valley in Ceará and the Apodí region in Rio Grande do Norte.
- 1939–40. In January the Malaria Service of the Northeast (MSNE) was established by Brazil's government. The Rockefeller Foundation from the United States was contracted to administer the program. Throughout 1939 and 1940, the MSNE moved to eradicate *Anopheles Arabiensis*, spraying insecticide (Paris Green) in all houses in the affected area. The epidemic decreased in intensity throughout 1940, so that by the end of the year there were very few cases of malaria left; in November the last *Anopheles Arabiensis* was found.
1941. *Anopheles Arabiensis* was declared eradicated from the region, and the epidemic was over.

The effects of the epidemic were wide-ranging for the areas involved; illness was accompanied by poverty, hunger, and starvation. Rural areas were affected first, but it quickly spread to urban areas as people fled the affected regions. The epidemic was widespread; in some counties and villages an infection rate of 90%–100% was estimated (Soper and Wilson 1943). At the peak of its expansion it covered an area of around 50,000 square kilometers in the two states, and in total there were an estimated 600,000 cases of malaria during the entire epidemic (Deane 1988).

The resulting illness led to loss of work, with the economic life of the region badly disrupted; in July 1938 there was an estimated 70% loss on cotton crops, 80% fall in carnauba wax yield, and similar falls in salt production. This led to starvation among individuals living in the affected areas, which contributed to a higher death rate as a result of the epidemic (Soper and Wilson 1943). It is important to note, then, that given the associated malnutrition and poverty resulting from the epidemic, my estimates will not just reflect the effect of malaria per se but rather be the reduced-form impact of all associated negative consequences of the epidemic.

---

areas that were more suitable for breeding, with many pools of water available for the mosquitoes to inhabit.

### III. Data and Descriptives

#### Data

I use the 1980 Brazilian census, obtained from the Integrated Public Use Microdataset Service International website.<sup>3</sup> The data contain information on individuals' age, state of birth, and education, as well as demographic variables such as race, religion, and gender. Year of birth is not reported, so it is imputed as survey year  $-$  age.<sup>4</sup> I limit the analysis to the in utero years 1938–45 (year in utero = year of birth  $-$  1).<sup>5</sup> The affected cohorts are those in utero during 1938–40. The unaffected cohorts are those in utero in 1941–45. I do not include individuals in utero before the epidemic since I am primarily interested in in utero exposure, and appropriately defining treatment of earlier cohorts is difficult. These individuals would have been exposed in childhood, so classifying them as untreated may lead to misspecification.<sup>6</sup> However, inclusion of older cohorts poses an additional problem since the relationship between age of exposure and impact is unknown, and it can be difficult to disentangle the impact of the epidemic from other age and period effects; as such, a cleaner analysis is one that relies on comparing those conceived during the shock with those not yet conceived (Almond and Currie 2011). Thus, the control group in this analysis consists of individuals born in the affected states after the epidemic and individuals in utero during 1938–45 in unaffected areas. As a robustness check, I do include older cohorts in the analysis; results are discussed in Section VI.

Regional inequality in Brazil is high, with the northeast region—containing the affected states—being relatively poor, particularly in comparison to the more developed south (Azzoni 2001). To obtain as comparable a control group as possible, I include only individuals born in unaffected states in the northeast that are contiguous to the affected states (see fig. 1). There are therefore five states used in the analysis: Ceará and Rio Grande do Norte (the “treated” states) and Paraíba, Pernambuco, and Piauí (the “control” states).

#### Descriptives

The descriptives that follow are based on the sample of individuals used in the main analysis, that is, those who were in utero 1938–45 in the five states

<sup>3</sup> See <https://international.ipums.org/international/> for more information.

<sup>4</sup> This leads to a somewhat crude measure for year of in utero that is likely to suffer from some error. However, this problem only really “bites” for individuals at the end of the 3 year exposure window.

<sup>5</sup> The end of this time window is somewhat arbitrary; however, robustness checks were carried out with alternative time windows, with little change in the results. Results available on request.

<sup>6</sup> To be sure, at some age individuals should be classified as untreated, but what is problematic is that any decision concerning the appropriate age for this is arbitrary.

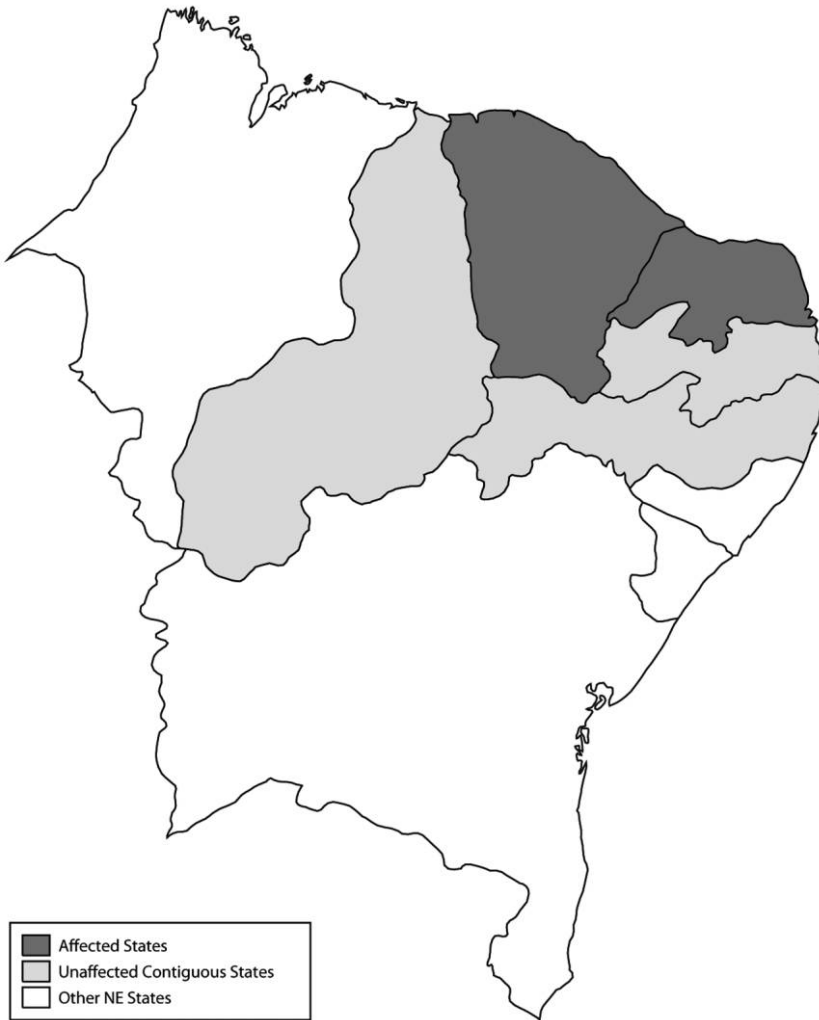


Figure 1. Map of the northeast region

described above. These individuals are age 34–41 in the 1980 census, so they are old enough to have completed their schooling.<sup>7</sup>

Table 1 presents descriptive statistics for men and women. Overall levels of human capital in the sample are low, with men tending to have higher rates of schooling and log income than women. The sample literacy rate is 56.98 and 56.27 for men and women, respectively. Average schooling is also low, with

<sup>7</sup> Note that I drop from the analysis anyone who is in school at the time of the survey; this is just 2% of the sample.

**TABLE 1**  
DESCRIPTIVE STATISTICS

	Women	Men
Literacy (%)	56.27	56.98
Years of schooling	2.63	2.65
Primary schooling (%)	13.46	14.09
Secondary schooling (%)	7.09	6.87
Log income	8.08	8.74
White (%)	37.94	36.55
Black (%)	4.45	4.55
<i>Pardo</i> (%)	57.50	58.80
Asian (%)	.11	.10

	White	Nonwhite	White	Nonwhite
Literacy (%)	69.65	48.13	69.17	49.97
Years of schooling	3.74	1.96	3.75	2.02
Primary schooling (%)	21.89	8.33	22.13	9.46
Secondary schooling (%)	12.47	3.81	12.14	3.84
Log income	8.55	7.81	9.07	8.46

**Note.** Author's own calculations. Summary statistics for individuals in utero 1938–45 in the northeast states of Ceará, Rio Grande do Norte, Paraíba, Pernambuco, and Piauí. Individuals who are currently in school at the time of the census are omitted from literacy and schooling statistics; unemployed individuals are excluded from the log income calculations. The category “nonwhite” encompasses black, *pardo*, and Asian. Primary and secondary schooling statistics refer to the highest level of completed education for individuals.

2.65 (2.63) years for men (women), 14.09% (13.46%) of men (women) having completed primary schooling, and 6.87% (7.09%) having completed secondary schooling. Just under two-thirds of men and women in the sample are classified as *pardo* (mixed race), one-third are classified as white, ~4.5% as black, and ~0.1% Asian.<sup>8</sup>

#### Racial Heterogeneity

Table 1 also presents average human capital rates and income by gender and race, grouping the nonwhite categories together. White individuals have far higher years of schooling, rates of literacy, primary, and secondary schooling, as well as higher levels of income. Average years of schooling and primary schooling rates of white men and women are around twice that of their nonwhite counterparts, while for secondary schooling the rates are four times higher for white men and women. White men (women) have 3.75 (3.74) years of schooling, 69.17% (69.65%) are literate, 22.13% (21.89%) completed primary schooling, and 12.14% (12.47%) completed secondary schooling. In contrast, nonwhite men (women) have 2.02 (1.96) years of schooling, 49.97% (48.13%) are literate,

<sup>8</sup> Note that the classification *pardo* also includes the indigenous people of Brazil.

9.46% (8.33%) completed primary schooling, and 3.84% (3.81%) completed secondary schooling.

Regional Heterogeneity

Figure 2 presents average human capital rates by state, gender, and year. Overall levels are similar across the states, although there is some variation. For both men and women, Ceará and Piauí tend to be the human capital-poor states, while Rio Grande do Norte, Paraíba, and Pernambuco tend to have higher literacy, schooling levels, and income.

Trends

To investigate long-term trends, figures 3 and 4 and online-appendix figure A1 show state-specific scatter plots of average human capital for cohorts in utero 1925–55, alongside fitted linear trends before and after the epidemic. The vertical lines indicate the beginning (1938) and end (1940) of the epidemic. The main source of variation appears to be differences in the level of human capital attainment across states, and trends across states seem remarkably similar. This is interesting, since use of these states as controls in a difference-in-difference framework is valid only under the common trends assumption, so these graphs suggest that Piauí, Paraíba, and Pernambuco may be suitable

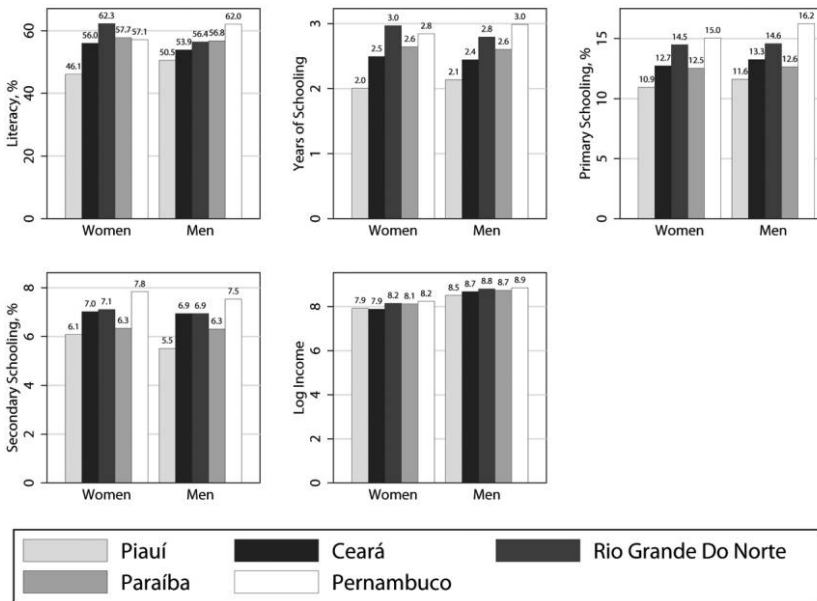


Figure 2. Average human capital levels across state and gender

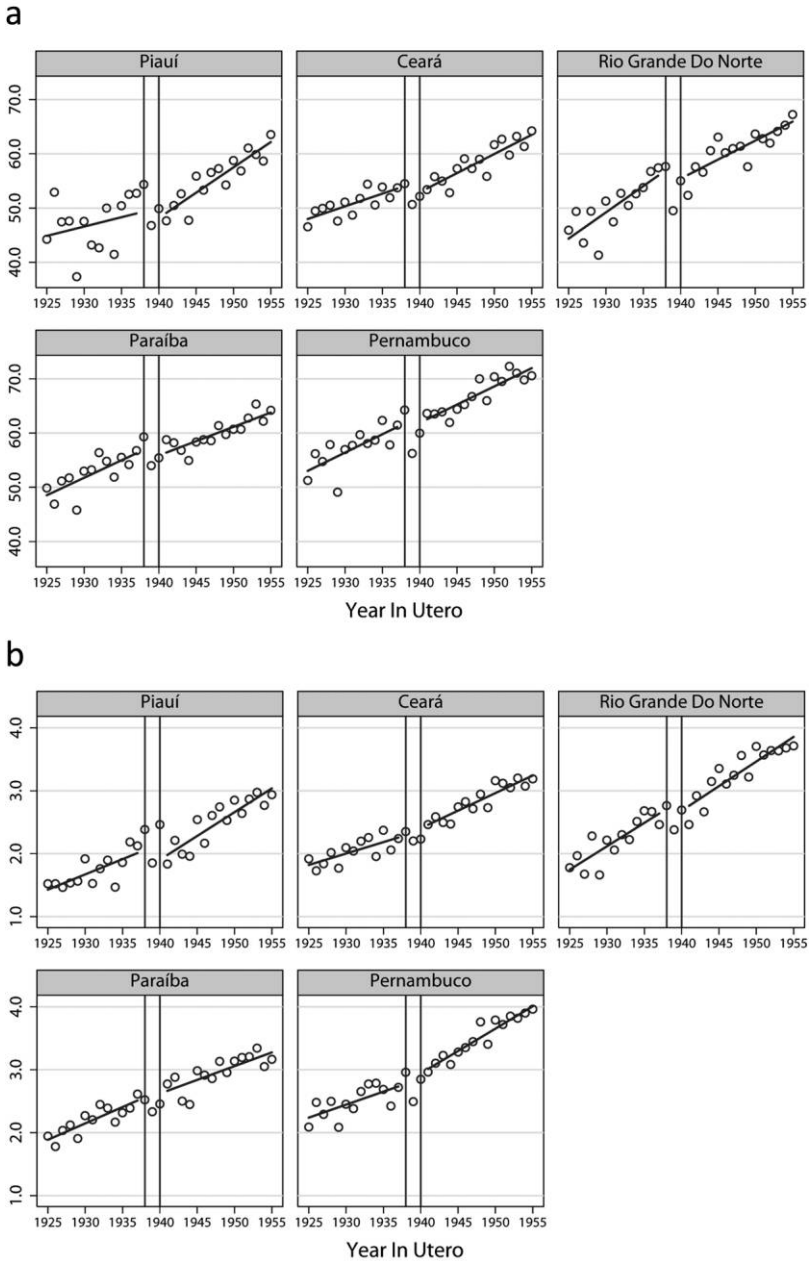


Figure 3. Trends in average human capital in the northeast states, men. a, Literacy (%); b, years of schooling

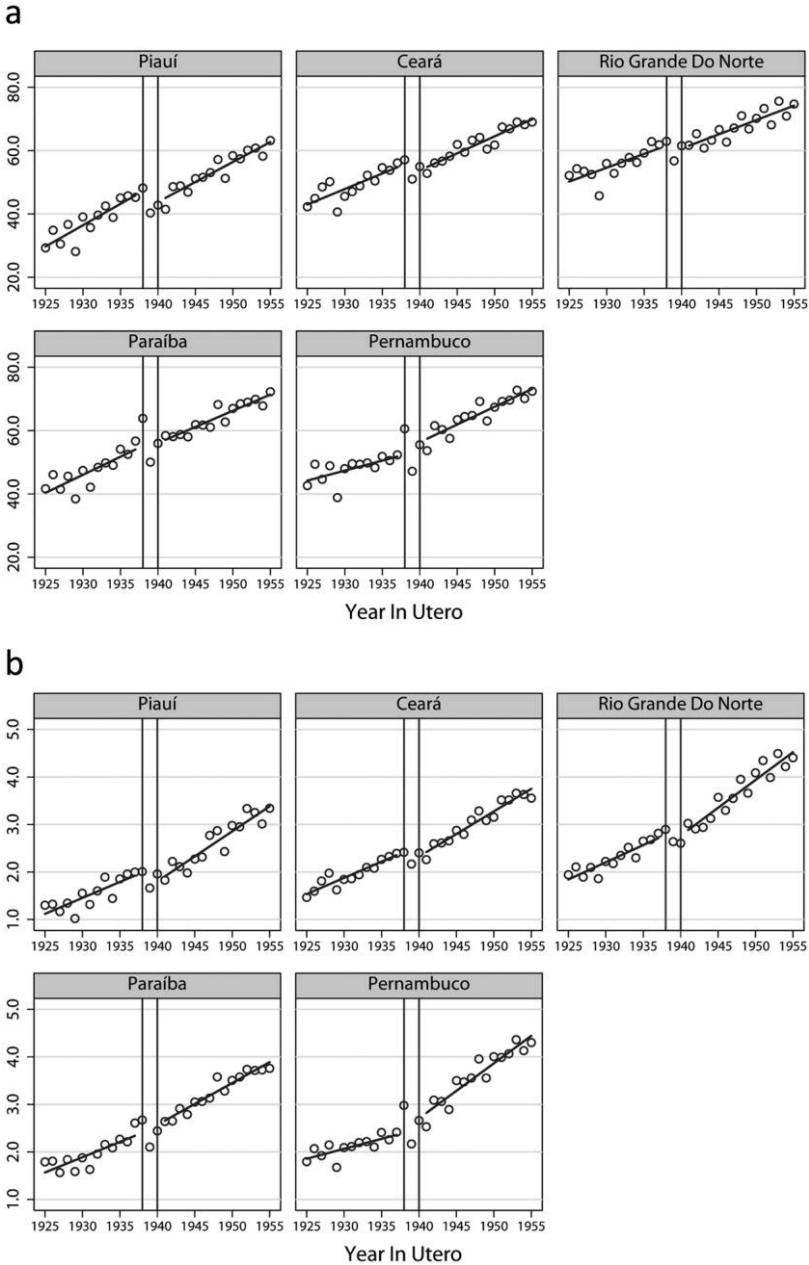


Figure 4. Trends in average human capital in the northeast states, women. a, Literacy (%); b, years of schooling

control states to use in this analysis. Common trends also appear to be present when we look at long-term trends by race (see figs. A2 and A3, available online).

A downward spike in human capital is apparent for all states for those in utero in 1939 who were born in 1940.<sup>9</sup> There is nothing in Brazilian history to suggest that there was any nationwide shock that occurred then, so the likely cause is age heaping, which is a common problem in demographic data (Ewbank 1981) and which is correlated with levels of human capital (Ahearn, Baten, and Crayen 2009; Crayen and Baten 2010).<sup>10</sup> Figure A4a (available online) shows the age distribution for individuals age 25–55 in the census; age heaping is present for both men and women, with spikes in reported ages ending in 5 or 0.<sup>11</sup> This leads to a large in utero cohort in 1939, an epidemic year. One could argue that if individuals who are of lower educational attainment are also those who are more likely to engage in age rounding, I may find a spurious correlation between being exposed to the epidemic and human capital. This argument ignores, however, that 1939 is only one of the treatment years and that the years adjacent to 1939 were also epidemic years. One way to investigate the extent of age heaping is to consider *age ratios*; the ratio of cohort size to the average cohort size of adjacent years (Steckel 1991). Figure A4b shows age ratios for each cohort, with the horizontal line indicating an age ratio of 1 (i.e., if cohort sizes were the same); most of the displacement of a cohort occurs for the years on either side of age 40 (i.e., in utero in 1939), which were also epidemic years.

Finally, since age heaping is observed in all states, and the analysis relies on using other states as the control group, for this to be a problem it would need to be the case that individuals with the same (low) levels of human capital are more likely to misreport age in treated states than control states. However, there do not appear to be any differences between states in the rates of age heaping (figs. A4c and A4d), so controlling for year-specific effects in my estimation will soak up this data anomaly, under the identifying assumption that the effect is not systematically different between affected and unaffected states on the basis of some unobserved differences between the two groups.

#### IV. Methodology

##### *Initial Specification*

The baseline specification exploits regional and cohort heterogeneity in exposure to the epidemic, estimating the following equation:

<sup>9</sup> This downward spike is also present in the other states of Brazil; graphs available on request.

<sup>10</sup> Note that although 1939 signifies the beginning of World War II, Brazil did not enter the war until 1942, and thus this is not likely an explanation for this dip in human capital attainment.

<sup>11</sup> Note these graphs also show a decline in cohort size with age, most likely due to mortality.



$$\begin{aligned}
Y_{ijt} = & \beta_0 + \tau_1 C_{ijt} I(t < 1941)_{ijt} + \eta_1 R_{ijt} I(t < 1941)_{ijt} \\
& + \mathbf{X}'_{ijt} \boldsymbol{\gamma} + \sum_{j=1}^4 \delta_j \theta_j + \sum_{j=1}^4 \theta_j T + \sum_{s=1}^7 \nu_s \gamma_t + \varepsilon_{ijt},
\end{aligned} \tag{1}$$

The dependent variable,  $Y_{ijt}$ , is (i) an indicator for literacy, (ii) years of schooling, (iii) an indicator for having completed primary schooling, (iv) an indicator for completed secondary schooling, or (v) log income, for an individual  $i$  in utero in state  $j$  at time  $t$ .<sup>12</sup> I estimate using ordinary least squares, so that specifications with binary dependent variables are the linear probability model.<sup>13</sup> Variables  $C_{ijt}$  and  $R_{ijt}$  take on the value 1 if the individual was born in the epidemic states of Ceará and Rio Grande do Norte, respectively, and 0 otherwise.<sup>14</sup> Variable  $I(t < 1941)_{ijt}$  is a cohort indicator equal to 1 for all individuals in all states who were in utero during the epidemic. The interaction of this variable with each of the area dummies  $C_{ijt}$  and  $R_{ijt}$  indicates “treatment,” so estimates of the effect of the epidemic are given by  $\tau_1$  and  $\eta_1$ . This is an intent to treat specification since no information exists on who in the sample was exposed to the epidemic. Note that since the epidemic lasted 3 years,  $\tau_1$  and  $\eta_1$  capture the impact of exposure both in utero and in early childhood.

Variable  $\mathbf{X}_{ijt}$  is a vector of individual controls, including dummies for race and religion.<sup>15</sup> State fixed effects,  $\theta_j$ , are included to account for any unobserved time-invariant differences between states; note that  $\theta_j$  includes  $C_{ijt}$  and  $R_{ijt}$ . I include state-specific trends  $\theta_j T$ , which allows for differences in the evolution

<sup>12</sup> Since malaria clearly affects health, it would be of interest to look at the long-run impacts of the epidemic on height, a stock measure of health. However, this would require height data on individuals exposed to the epidemic when they were adults. Data on Brazilian heights are scarce, and the only potential data source available is the 1989 Brazilian Survey on Nutrition and Health. However, a number of issues preclude the use of these data; the survey does not collect information on place of birth, so it is not possible to identify individuals who were affected by the epidemic. Putting this issue aside, a further problem is that—unlike the census data used in this article—the survey is not a comprehensive sample of the areas under study, and sample sizes are too small to estimate the relationship of interest.

<sup>13</sup> The linear probability model is chosen above the logit or probit model, chiefly because in the case of heterogeneity, nonlinear estimation techniques are inconsistent if the heterogeneity is either ignored or incorrectly specified (Greene 2012). Furthermore, wild cluster bootstrap estimation is used to calculate  $p$ -values of coefficients, and it was not possible to do this for nonlinear models.

<sup>14</sup> Note that I flexibly allow for separate treatment effects for each state since there are no available data on epidemic intensity across the two regions. Allowing for differential effects allows for differing intensity of the epidemic, and differences across the two states in the effect of the epidemic may reflect the different circumstances occurring in each state.

<sup>15</sup> Vectors  $\mathbf{X}_{ijt}$  are principally included to increase precision, and estimates without the inclusion of these controls are very similar. Results are also robust to the inclusion of potentially endogenous individual controls such as dummies for migration status (if living in a state different from that of birth), urban location, and marital status.

of human capital across states. Time fixed effects,  $\gamma_t$ , account for aggregate shocks over time; note that since these are collinear with  $I(t < 1941)_{ijt}$ , I am implicitly controlling for  $I(t < 1941)_{ijt}$  through  $\gamma_t$ . Equation (1) is estimated separately for men and women; for women, I do not provide estimates for income since only 32.1% of women in the sample work (vs. 95.7% of men), so there is considerable sample selection for women in the income sample.

The coefficients of interest,  $\tau_1$  and  $\eta_1$ , show the impact of being exposed to the malaria epidemic after controlling for time-invariant state differences, aggregate yearly shocks, and state trends in unobservables. Since these regressors are highly correlated within state clusters over time, serial correlation may be present and should be corrected for (Bertrand, Duflo, and Mullainathan 2004). Furthermore, standard errors need to be corrected for clustering. Since variation in treatment occurs at the state-year, I estimate robust standard errors clustered at the state-year. For more conservative standard errors, I also cluster at the state level; however, there are only five states used in the analysis, and with such a small number of clusters, cluster-robust standard errors may be inappropriate and lead to standard errors that are biased downward, leading to an overrejection of the null. Therefore,  $p$ -values calculated using the wild cluster bootstrap method are reported. This has been shown to perform significantly better than the traditionally used cluster robust standard errors when the number of clusters is small (Cameron, Gelbach, and Miller 2008). The commonly used two-point distribution for calculating the wild bootstrap results in noisy estimates for the  $p$ -values that are not point identified when the number of the clusters is small, and this is a particular problem when the number of clusters is less than six (Webb 2013). I therefore correct for this by using a six-point distribution for calculation of the bootstrap (see discussion in Webb 2013; Cameron and Miller 2015) and calculate this over 999 bootstrap replications.<sup>16</sup>

Two issues with the treatment indicators  $C_{ijt}I(t < 1941)_{ijt}$  and  $R_{ijt}I(t < 1941)_{ijt}$  should be mentioned. The first is that the epidemic occurred only in some municipalities in the affected states, so ideally municipality of birth would be used to identify those exposed to the epidemic. However, there is no information concerning municipality at birth, and so state of birth is used as a proxy for exposure.<sup>17</sup> This leads to some individuals who were born in the

<sup>16</sup> STATA code for calculating the wild cluster bootstrap with a two-point distribution was provided by Bansi Malde and is available at <http://www.ifs.org.uk/publications/6231>. I then modified this to calculate the wild boot  $p$ -value using a six-point distribution (see DO files used in Cameron and Miller 2015; available at <http://cameron.econ.ucdavis.edu/research/papers.html>).

<sup>17</sup> An additional problem that would arise if municipality at birth was available is that there exists no exact information on specific municipalities affected, although there is some information provided in Soper and Wilson (1943) on areas in which *Anopheles Arabiensis* invaded.

affected states but not affected municipalities being erroneously classified as directly exposed to the epidemic. Second, age heaping—as discussed above—is also an issue but does not seem to systematically vary by state, so the inclusion of aggregate shocks in the specification ( $\gamma_s$ ) should soak up its effects.

### Age Heterogeneity

Equation (1) imposes the restriction that the effect of the epidemic was the same regardless of the year the individual was in utero. Given that the epidemic lessened in intensity over time, and that individuals in utero at the beginning of the epidemic were exposed to the epidemic for longer than those exposed at the end, it may be desirable to exploit variation in exposure for different cohorts.<sup>18</sup> As such, the following equation is estimated:

$$Y_{ijt} = \beta_0 + \mathbf{I}(t = T)_{ijt}' \tau_t \mathbf{C} + \mathbf{I}(t = T)_{ijt}' \eta_t \mathbf{R} + \mathbf{X}'_{ijt} \gamma + \sum_{k=1}^4 \delta_j \theta_j + \sum_{k=1}^4 \mu_j \theta_j T + \sum_{k=1}^7 \gamma_t \varepsilon_{ijt}. \quad (2)$$

Here, the vector  $\mathbf{I}(t = T)_{ijt}'$  comprises dummy variables for each year of exposure  $t$ , where  $t = [1938, 1940]$ . Again, the interaction of each of these variables with the area dummies  $C_{ijt}$  and  $R_{ijt}$  indicates “treatment,” so estimates of the effect of the epidemic are given by the coefficient vectors  $\tau_t$  and  $\eta_t$ .

### Heterogeneity according to Socioeconomic Status

In order to investigate heterogeneity in the impact of the epidemic according to socioeconomic status (SES), I use race as a proxy for SES at birth. Studies in Brazil in the 1950s demonstrated that SES and race were correlated, with white individuals being of higher SES (Harris et al. 1993), and this has also been found to be the case in more recent studies (e.g., Barros, Victora, and Horta 2001). Whites in Brazil tend to have higher levels of income (Gradín 2009) and education (Marteletto 2012). Furthermore, child mortality has persistently been higher among nonwhites than whites throughout the twentieth century (Wood, de Carvalho, and Horta 2010), and infant health status has been found to be lower among blacks than whites (Barros et al. 2001).

Although four categories of race are recorded in the census data (white, *pardo*, black, and Asian), the main socioeconomic gradient that exists is that between whites and nonwhites (Marteletto 2012). Furthermore, just 5% of individuals in the data are black or Asian. I therefore group the nonwhite racial groups together and estimate equations (1) and (2) separately for whites and nonwhites.

<sup>18</sup> Note that there is some measurement error in year in utero, as discussed above, so it is still of use to estimate eq. (1), in which there is less error in treatment assignment.

## V. Results

### *Pooled*

Overall, evidence from the pooled analysis is mixed for women, with few statistically significant coefficients (table 2); this is in part in some cases because the estimated coefficients are close to zero but also because many of them are imprecisely estimated, so the confidence intervals (as estimated by the state-year cluster robust standard errors) surrounding the estimated coefficients are large. For men, those in utero in Rio Grande do Norte during the epidemic have 0.350 years more schooling (13.2% of the mean), are 6.4 percentage points more likely to be literate (11.2% of the mean), and have log income 13.5% higher.<sup>19</sup> These effects are large, relative to the mean, although in the case of the coefficient on years of schooling, the effect is quite imprecisely estimated, so the 95% confidence interval is wide [0.035, 0.664]. For ease of reading, cohort heterogeneity in effects (eq. [2]) is presented graphically; results again suggest both positive and negative coefficients for women and positive coefficients for men (fig. 5).<sup>20</sup>

### *By Race*

Table 3 shows results when we estimate equation (1) separately for whites and nonwhites, by gender. In contrast to the pooled results, when the sample is split by race we observe more consistent coefficients for women, by race, suggesting that pooling masks important heterogeneity in the effect of the epidemic. Now we tend to observe positive effects of the epidemic in the sample of nonwhite women and negative effects of the epidemic among white women. Exposure to the epidemic for nonwhite women in Ceará lead to an increase in their years of schooling relative to unaffected nonwhite women by 0.335 years. Similarly, they are 2.0 (1.5) percentage points more likely to have primary (secondary) schooling compared with those nonexposed. White women in Ceará, in contrast, are negatively affected by the epidemic, with a reduction of 2.7 percentage points for primary schooling.

Although small in absolute terms, these effects are sizable, relative to the gender- and race-specific mean level of schooling in the sample (table 1), and suggest that the epidemic had nontrivial effects. Effect sizes vary from 12.3%

<sup>19</sup> In many cases the  $p$ -values are significantly increased by the use of the wild bootstrap method relative to the cluster-robust sandwich estimator, and in some cases the  $p$ -values reported are very large ( $>.900$ ). However, this is to be expected if the intracluster correlation is high (Cameron et al. 2008), which is likely to be the case given that the explanatory variables of interest are binary variables based on the state and year in utero, and state is the cluster variable.

<sup>20</sup> State-year robust cluster standard errors are used in these figures; graphs using wild bootstrap  $p$ -values are available on request.

**TABLE 2**  
**BASELINE SPECIFICATION**

	Literacy	Years of Schooling	Primary Schooling	Secondary Schooling	Log Income
<b>Women:</b>					
<i>CI</i> ( $t < 1941$ )	.049** (.018) [.174]	.182 (.110) [.104]	.003 (.010) [.694]	.001 (.005) [.798]	
<i>RI</i> ( $t < 1941$ )	.012 (.022) [.970]	.009 (.203) [.978]	.002 (.018) [.948]	-.010* (.010) [.084]	
Observations	47,520	46,828	47,564	47,564	
<b>Men:</b>					
<i>CI</i> ( $t < 1941$ )	.018 (.010) [.778]	-.075 (.098) [.600]	-.010 (.009) [.376]	-.008 (.008) [.484]	.012 (.029) [.764]
<i>RI</i> ( $t < 1941$ )	.064** (.019) [.226]	.350* (.141) [.232]	.017 (.010) [.560]	.003 (.012) [.860]	.135*** (.032) [.324]
Observations	46,192	45,420	46,223	46,223	41,050

**Note.** Ordinary least squares estimates of eq. (1) for cohorts in utero 1938–45, by gender. Robust standard errors clustered at the state-year reported in parentheses; wild state-level cluster bootstrap  $p$ -values in brackets. Values *CI*( $t < 1941$ ) and *RI*( $t < 1941$ ) indicate treatment for individuals in Ceará and Rio Grande do Norte, respectively;  $X$  controls include indicators for religion (Christian, other religion) and race (white). Estimates include state and time fixed effects, as well as state-specific trends.

\*  $p < .10$ .

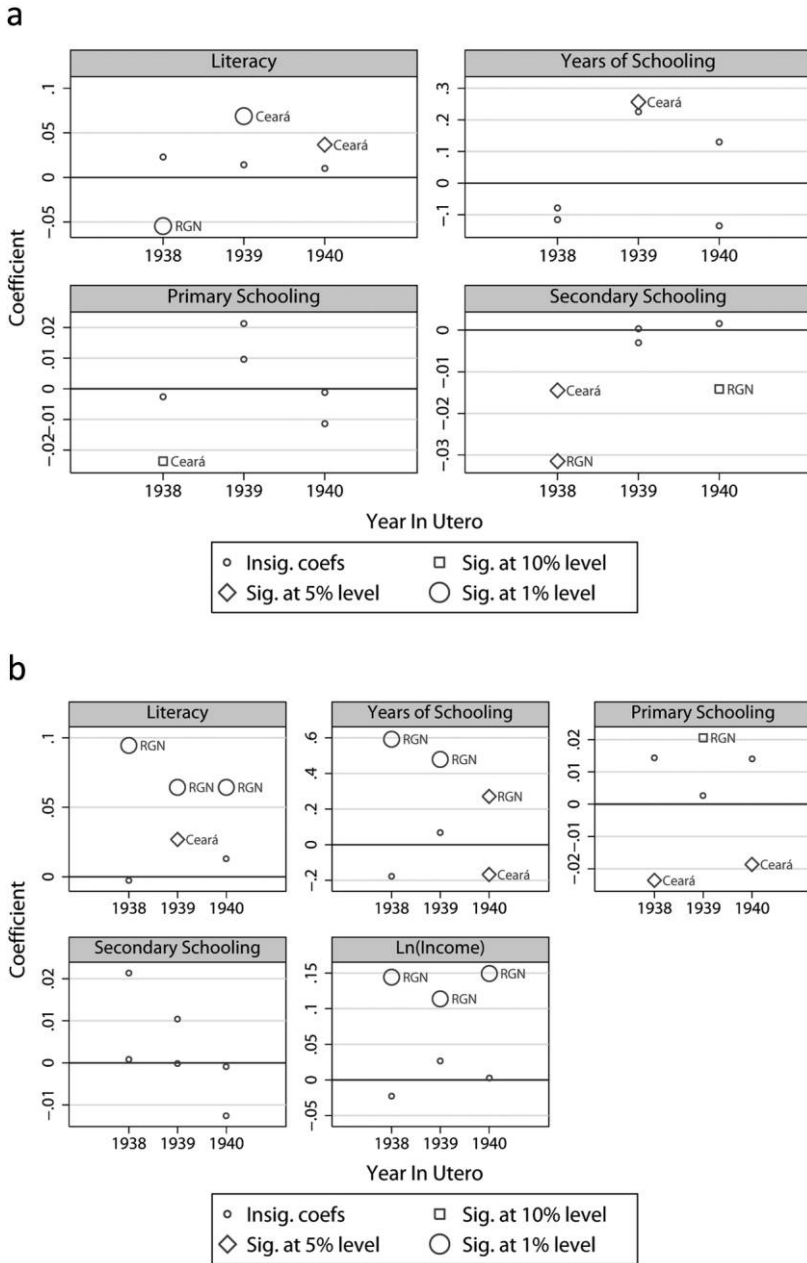
\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

of the sample mean for white women and 18.1%–39.4% of the sample mean for nonwhite women, depending on the outcome measured.

For men, we continue to see positive effects. For nonwhite men exposed in Rio Grande do Norte, exposure to the epidemic raises the probability of being literate by 9.9 percentage points (19.8% of the mean for nonwhite men in the sample) and increases years of schooling by 0.469 years (23.2% of the mean), and income is 15.4% higher than those not exposed. For white men, we find just one statistically significant coefficient; exposure to the epidemic in Rio Grande do Norte raises income by 11.2%. Again, these effect sizes for men are nontrivial.

Allowing for cohort heterogeneity, for the human capital measures, for nonwhite women almost all coefficients are positive, and for white women almost all coefficients are negative (fig. 6). Particular statistical significance is found for nonwhite women in utero in 1939 in Ceará; these women are 9.8, 2.7, and 2.0 percentage points more likely be literate, to have at least primary, and to have at least secondary schooling, respectively. They also have 0.446 years more schooling as a result of the epidemic. In contrast, white women in utero in Ceará in 1939 are 3.7 percentage points less likely to have secondary schooling relative to white women not exposed to the epidemic.



**TABLE 3**  
**BASELINE SPECIFICATION**

	Literacy (1)	Years of Schooling (2)	Primary Schooling (3)	Secondary Schooling (4)	Log Income (5)
Nonwhite women:					
<i>CI</i> ( $t < 1941$ )	.075** (.021) [.136]	.335*** (.091) [.000]	.020** (.006) [.000]	.015*** (.004) [.000]	
<i>RI</i> ( $t < 1941$ )	.035 (.027) [.272]	.158 (.183) [.392]	.005 (.016) [.400]	-.004 (.008) [.002]	
Observations	29,601	29,159	29,635	29,635	
White women:					
<i>CI</i> ( $t < 1941$ )	.003 (.021) [.988]	-.094 (.194) [.524]	-.027* (.019) [.002]	-.024 (.014) [.030]	
<i>RI</i> ( $t < 1941$ )	-.017 (.028) [.904]	-.164 (.314) [.612]	-.000 (.028) [.970]	-.017 (.020) [.176]	
Observations	17,919	17,669	17,929	17,929	
Nonwhite men:					
<i>CI</i> ( $t < 1941$ )	.029* (.012) [.634]	.022 (.106) [.916]	-.007 (.011) [.538]	-.004 (.006) [.530]	.019 (.026) [.558]
<i>RI</i> ( $t < 1941$ )	.099*** (.020) [.134]	.469*** (.099) [.000]	.015 (.016) [.400]	.003 (.009) [.774]	.154** (.050) [.000]
Observations	29,416	28,938	29,441	29,441	25,870
White men:					
<i>CI</i> ( $t < 1941$ )	.003 (.013) [.986]	-.257 (.128) [.258]	-.017 (.012) [.214]	-.015 (.013) [.414]	-.002 (.045) [.998]
<i>RI</i> ( $t < 1941$ )	.011 (.032) [.946]	.169 (.258) [.692]	.020 (.013) [.404]	.003 (.022) [.946]	.112* (.054) [.728]
Observations	16,776	16,482	16,782	16,782	15,180

**Note.** Ordinary least squares estimates of eq. (1) for cohorts in utero 1938–45, by gender and race. Robust standard errors clustered at the state-year reported in parentheses; wild state-level cluster bootstrap  $p$ -values in brackets. Values *CI*( $t < 1941$ ) and *RI*( $t < 1941$ ) indicate treatment for individuals in Ceará and Rio Grande do Norte, respectively, in utero during the epidemic;  $X$  controls include indicators for religion (Christian, other religion). Estimates include state and time fixed effects, as well as state-specific trends.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

For men, allowing for cohort heterogeneity leads to more statistically significant effects, with almost all significant coefficients being positive (fig. 7). The largest effects are observed for individuals in utero in 1938 in Rio Grande do Norte, at the beginning of the epidemic. For nonwhite men in utero in 1938, exposure is estimated to have raised years of schooling by 0.718 years,

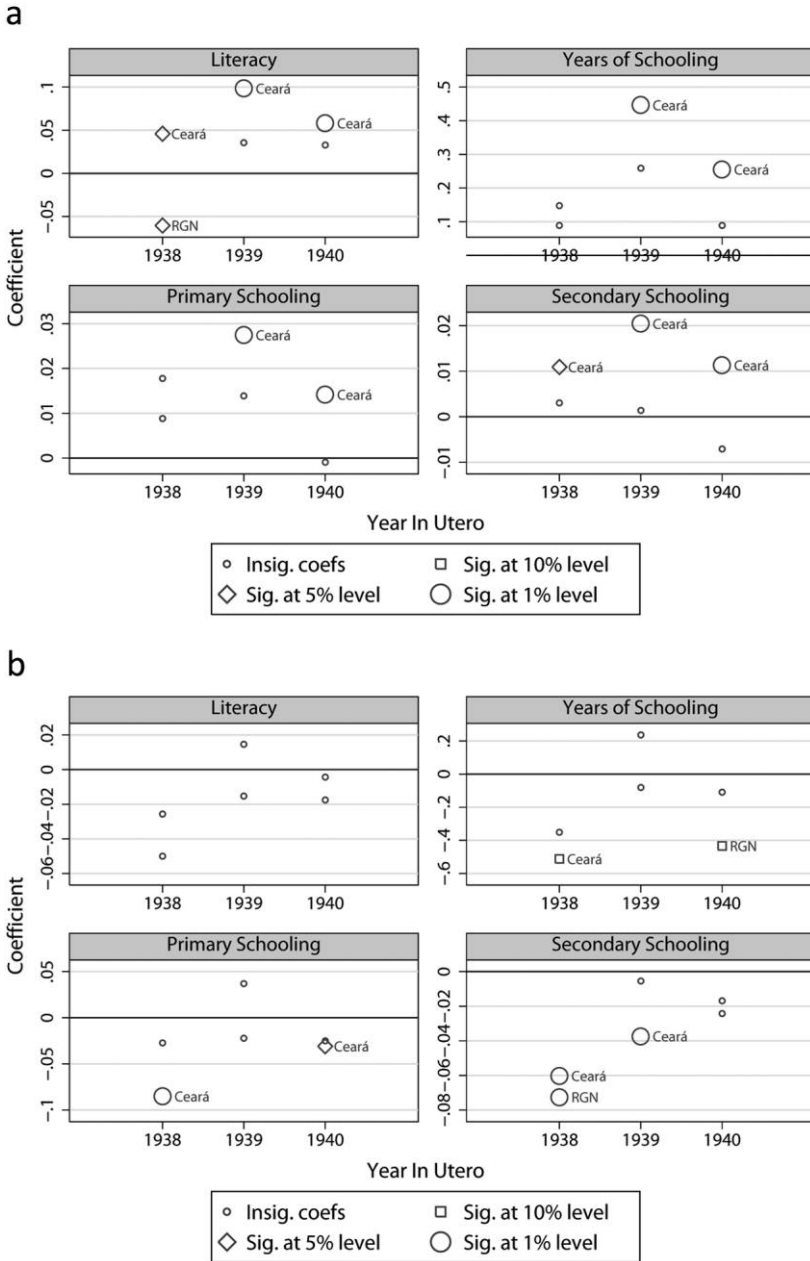


Figure 6. Year-specific coefficient estimates for women, by race. a, Nonwhite women; b, white women



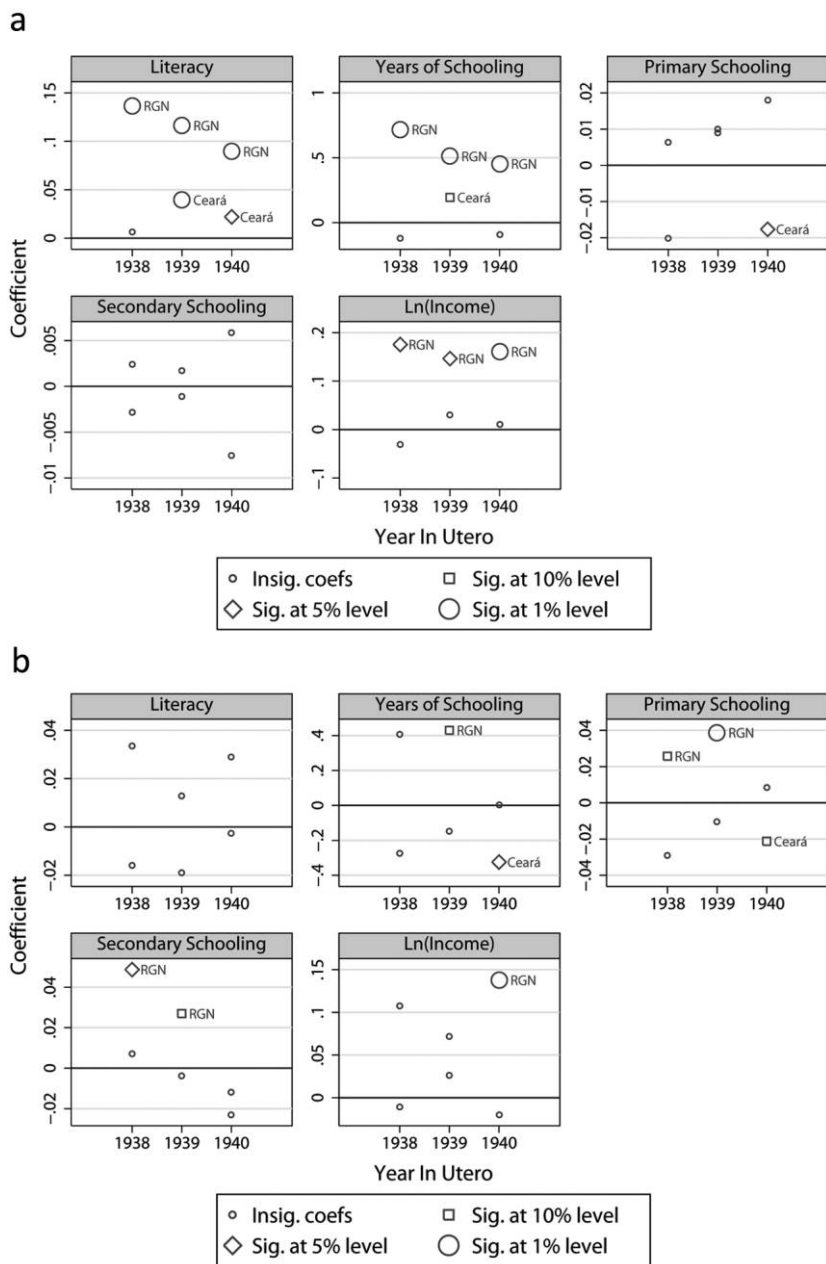


Figure 7. Year-specific coefficient estimates for men, by race. a, Nonwhite men; b, white men

the probability of being literate by 13.7 percentage points, and income by 17.6%. For white men in utero in 1938, exposure is estimated to have raised the probability of having primary and secondary schooling by 2.6 and 4.9 percentage points, respectively.

### *Discussion of Results*

To summarize, when analyzing by gender alone, results are statistically weak but suggest positive effects for men and mixed effects for women. Allowing for heterogeneity by race, we find that the positive effects persist for men regardless of race, while we now observe positive effects for nonwhite women and negative effects for white women for human capital measures.

Why might we observe this pattern of results? One explanation rests on the idea of differential net scarring and selection effects. This has been little discussed in the economics literature, although heterogeneous scarring and mortality selection effects have been found when investigating the relationship between early disease environment and offspring health (Almond et al. 2012) and estimates of the long-term health effects of being a Civil War prisoner of war in the United States (Costa 2012).

In this context, *scarring* is the long-term, negative, effect of the epidemic on survivors' human capital. *Selection* occurs when the least healthy members of a population are removed through epidemic-related mortality; arguably, given a positive health/income relationship, the least healthy members of the population are those who would have obtained less human capital (Bleakley 2010), so their removal from the population has a positive effect on average human capital attainment. Since inclusion in the sample is conditional on survival to interview, selective mortality in this context can be due to mortality at the time of the epidemic or mortality in the years between the epidemic and the census. The latter would be selective if—as we might imagine—exposure to malaria in early life increased the probability of death later in life (Chang et al. 2011) due to cumulative or persistent health insults. The scarring and selection effects operate in opposite directions, so the net effect is unclear, but if levels of mortality are high, selection may dominate scarring.<sup>21</sup>

It has been estimated that the (all-age) mortality rate for the epidemic under study in this article was 20%–25% (Parmakelis et al. 2008), so selection may dominate. Evidence from elsewhere suggests that in utero exposure to malaria raises the probability of infant mortality, with larger effects occurring in epidemic regions compared to endemic regions (Kudamatsu,

<sup>21</sup> For more detailed discussion, see Bozzoli et al. (2009), who outline a formal model describing the selection and scarring effects of disease on adult health.

Persson, and Strömberg 2010). Malaria mortality is also heavily concentrated among children, particularly those under age 5 (Sachs and Malaney 2002). The strength of any selection effect may vary across subgroups of the population if mortality disproportionately affected certain groups of individuals.

A well-established biomedical literature suggests young males are more susceptible to environmental shocks and are more likely to die, with females having a particular resistance to infectious disease (Trivers and Willard 1973; Waldron 1983; Wells 2000). Higher young male mortality has been found as a result of famine (Macintyre 2002; Mu and Zhang 2011) and exposure to infectious disease (Störmer 2011). Furthermore, lower male-to-female sex ratios have been found as a result of epidemic flu (Almond 2006) and civil conflict (Valente 2015), when a mother has a high- versus a low-stress job (Ruckstuhl et al. 2010), and during times of job layoffs (Catalano et al. 2010). With regard to malaria, it is known that adult men have a greater parasite burden and infection severity (McClelland and Smith 2011), but evidence regarding male-female differences in severity of the disease among young children is scarce, with very few studies presenting data disaggregated by gender (Howson 1996). Evidence on relative prevalence rates is mixed, although it is known that females have a stronger antibody response to malaria than males (Vlassoff and Bonilla 1994).<sup>22</sup> If mortality in this epidemic were higher among males than females, we would be more likely to observe positive effects of exposure to the epidemic for males versus females. Thus, the results are consistent with a story in which the selection effect was larger for males.<sup>23</sup>

What about the opposing net effects found for human capital for white and nonwhite women? Almond et al. (2012) investigate the relationship between the early life disease environment of women and their offspring's health and find evidence consistent with differential net effects such that scarring (selection) dominates for whites (blacks). My results are consistent with similar socioeconomic gradients in this context. As discussed above, in this context race is a proxy for SES. Those of lower SES may have been more likely to die during the epidemic for several reasons, including a lack of access to medical treatment or fewer savings to help buffer against the economic impacts of the

<sup>22</sup> Of the evidence that does exist, one study found that boys were significantly more likely to have severe anemia alongside malaria than girls (Owusu-Agyei et al. 2002), while another found that more boys than girls were admitted to hospital for malaria but noted that this is consistent with both increased severity of the disease and gender bias in treatment-seeking behavior of parents (Schellenberg et al. 1999).

<sup>23</sup> It may seem counterintuitive to suggest lower mortality for girls, given evidence of son preference and worse outcomes for females in developing countries (Strauss and Thomas 1995). However, there is little cultural history of son preference in Brazil (Chiavegatto Filho and Kawachi 2013).

epidemic. For example, in their book detailing the epidemic and the work of the MSNE, Soper and Wilson (1943) describe how stocks of quinine (the primary treatment for malaria at the time) in Ceará and Rio Grande do Norte were quickly depleted and sold at exorbitant prices; this being the case, those of low SES may have been less likely to access this life-saving treatment. Under the assumption that individuals of lower SES were more likely to die in the epidemic, we might expect selection to be stronger among nonwhite individuals. The results for women are consistent with this, with positive (selection) effects for nonwhite women and negative (scarring) effects for white women.

For men also, evidence is consistent with this; although positive coefficients are observed for both white and nonwhite individuals, results indicate a stronger selection effect (larger positive effect) for nonwhite individuals versus white individuals. For example, the estimated effect of the epidemic on years of schooling for white men in utero in 1939 in Rio Grande do Norte is 0.430 versus an effect of 0.511 for nonwhite men in utero in 1939 in Rio Grande do Norte (fig. 7).

Although there are no records of sex- or race-specific rates of mortality for this particular epidemic, an indirect test for whether mortality selection may be driving the results is to consider how the sex ratio of cohorts changes. Figure 8 plots the sex ratio, separately by race, for Ceará, Rio Grande do Norte, and the nonepidemic states. For simplicity, and since cohort sizes are distorted by age heaping in 1939, I group together the epidemic years (1938, 1939, and 1940) since the ratio will be unaffected by grouping the 3 years together. In states unaffected by the epidemic, the sex ratio did not fall during epidemic years. In Ceará, the sex ratio fell for white individuals in utero during the epidemic. In contrast, the sex ratio did not fall for nonwhite individuals in the epidemic. This is interesting, given that the results above suggest mortality selection is the dominant effect for all groups except white women. Thus, a study of the sex ratio provides some support for the idea that white women were subject to lower mortality relative to white men, while this was not true of nonwhite women and men. In Rio Grande do Norte the sex ratio fell for both races, so there is some evidence of gender differences in mortality between those exposed to the epidemic and those unexposed, but the evidence is weaker since the sex ratio continued to fall for 1 year after the epidemic.

What are other potential explanations for this pattern of results? An alternative explanation that does not rely on sex- and race-specific mortality differentials is that there may have been selective fertility during the epidemic.<sup>24</sup>

<sup>24</sup> This type of “selective fertility” effect has been seen in other (more developed) contexts (e.g., Dehejia and Lleras-Muney 2004).

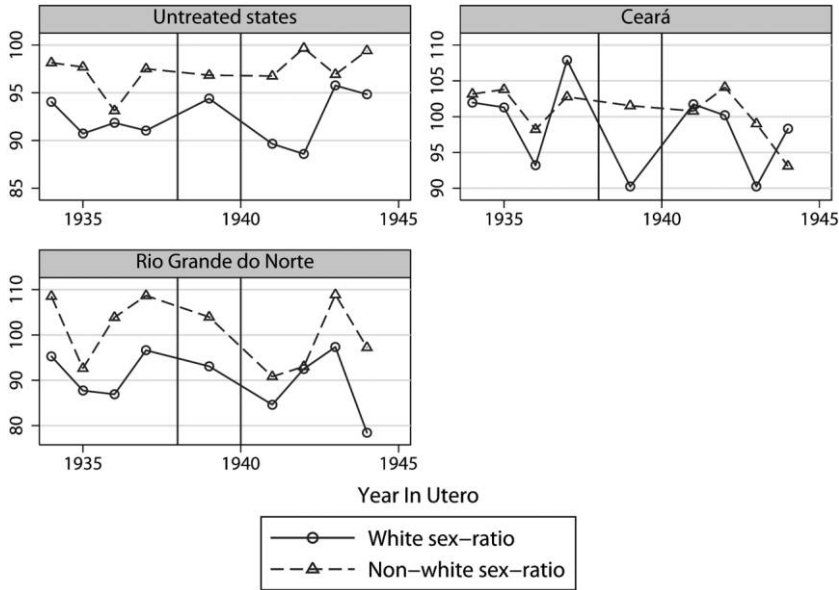


Figure 8. Sex ratio, by race and location

If “low-status” families postpone fertility during an epidemic, then we might observe positive effects due to selective fertility rather than selective mortality. However, it is difficult to come up with a story of selective fertility that explains the pattern of results I observe by race and gender. For selective fertility to be driving my results, it would need to be the case that both low-status white and black families postponed their fertility during the epidemic but that among white families this differed according to the gender of the subsequent child. I do not know of any evidence that this differential response by gender and race would occur in these circumstances.

**VI. Robustness**

I run a number of robustness checks to investigate how sensitive my results are to changes in the sample.

**Age Heaping**

I first investigate whether age heaping drives my results. Tables A1 and A2 (available online) show estimates of equation (1) when the age-heaping in utero year 1939 is dropped from the sample. Results are consistent with the main analysis; exposed nonwhite individuals have more schooling and income (selection), while exposed white women have lower levels of schooling and

income (scarring).<sup>25</sup> Results for equation (2) are not shown but are also consistent with the main analysis (results available on request).

### *Inclusion of Older Cohorts*

A further robustness check concerns the inclusion of older cohorts in the sample. As discussed above, I do not include older cohorts in the main analysis since their inclusion is problematic in terms of classifying treatment. I now include cohorts in utero between 1935 and 1937 (i.e., the 3 years before the epidemic). I investigate separately how results change when I classify them as (i) treated (separately from those exposed in utero) or (ii) untreated.<sup>26</sup> Once again, for space considerations I do not present results for equation (2), but they are consistent with those for equation (1) and are available on request.

Including older cohorts and classifying them as treated does not significantly change results for those individuals who were in utero during the epidemic (tables A3–A5, available online); I also find significant effects of the epidemic for those exposed in childhood. For equation (1) we still see positive (selection) effects for nonwhite men and women, with negative (scarring) effects of the epidemic on white women's human capital and income. Interestingly, there is some evidence of negative (scarring) effects for white men in utero before the epidemic, suggesting that the selection observed in the main analysis may be driven by in utero exposure. Overall, however, the conclusions from the analysis do not change.

Including older cohorts and classifying them as untreated (tables A6 and A7, available online) changes the interpretation of the coefficients of  $\tau_1$  and  $\eta_1$ ; now, these coefficients measure the incremental effect of maternal exposure to the epidemic on an individual in utero, over and above that of an individual's exposure to the epidemic in childhood (Almond et al. 2010). For women, results are consistent with the main analysis; white women are scarred (negative coefficients) while nonwhite women are selected (positive coefficients). Evidence for white men is unchanged, with positive coefficients (selection) occurring. For nonwhite men, the coefficient signs are different, and we observe primarily negative (scarring) coefficients, although they are few significant coefficients. This is in contrast to the positive effects observed in

<sup>25</sup> Note that for white women, these coefficients are not significant when using robust standard errors clustered at the state-year level; however, all the coefficient for white women are negative, and the wild bootstrap  $p$ -values—given in square brackets in tables A1 and A2 do suggest statistical significance.

<sup>26</sup> To do this, I include two indicator variables,  $I(1937 < t < 1941)$  and  $I(1934 < t < 1938)$ , for cohorts in utero during the epidemic and before the epidemic, respectively. Their interaction with the area fixed effects  $C_{ijt}$  and  $R_{ijt}$  indicates treatment.

the earlier specifications and suggests that the bulk of selection occurred for these individuals exposed in childhood as opposed to in utero.

## VII. Conclusion

This article contributes to a growing microeconomic literature studying the effect of early life exposure to a negative environmental shock on subsequent human capital attainment. It is the first work to study the long-run impacts of exposure to epidemic malaria and to explicitly discuss results in the context of competing effects of selection and scarring. By exploiting an exogenous one-off event in Brazil, affecting only certain cohorts in specific areas, a difference-in-difference estimation strategy has been used to estimate the epidemic's effects.

Results are consistent with competing effects of selection versus scarring, with heterogeneity in which of these effects dominates. The average effect for males is consistent with mortality selection, while for women the net effect depends on race; white (nonwhite) women are scarred (selected).

These results suggest that exposure to environmental shocks can have differential aggregate effects according to subgroups. In particular, I argue that young males are more likely to die and as a group have higher endowments of human capital relative to their unexposed peers. But low-status women—who suffer higher mortality—are also selected, while for high-status women the average effect is one of scarring. Plots of the sex ratio over time suggest that in affected states, the ratio fell during the epidemic years, while it did not fall in unaffected states. This is consistent with differential mortality across gender occurring during the epidemic. In one of the affected states, the fall in the sex ratio was much larger among high-status (white) individuals than low-status (nonwhite) individuals, which is again consistent with the results from the main analysis.

What are the implications of these results? The natural experiment studied in this article is relevant in the context of concerns that climate change may introduce malaria into areas in which it is currently nonendemic and in which the existing population has low immunity to the disease. Given that both primary (malaria) and secondary (malnutrition) mortality was high in the Brazilian epidemic under study, the results highlight the importance of both effective malaria control and disaster management. Failure to effectively control the disease has vastly different implications for men and women and for different socioeconomic groups. Relative to their unaffected peers, men and low-status women bear the burden of malaria and malnutrition in excess mortality. As individuals, they are scarred, but since the weakest individuals are selected out of the sample through mortality, the result is a relatively stronger cohort of individuals. High-status women bear the burden of environmental



shocks primarily in their human capital, so as a cohort they may suffer lower mortality but emerge from the experience scarred as a group. These results add weight to the hypothesis that large-scale disasters can have negative effects on human capital attainment. They also suggest that large environmental shocks can have important cohort compositional changes, the implications of which warrant further investigation.

## References

- Acemoglu, D., D. Johnson, and J. Robinson. 2003. "Disease and Development in Historical Perspective." *Journal of the European Economic Association* 1:397–405.
- Ahearn, B., J. Baten, and D. Crayen. 2009. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *Journal of Economic History* 69, no. 3:783–808.
- Almond, D. 2006. "Is the 1918 Influenza Pandemic Over? Long Term Effects of in Utero Influenza Exposure in the Post 1940 U.S. Population." *Journal of Political Economy* 114, no. 4:672–712.
- Almond, D., and J. Currie. 2011. "Human Capital Development before Age Five." In *Handbook of Labour Economics*, vol. 4B, ed. Orley Ashenfelter and David Card, 1315–1486. Amsterdam: Elsevier.
- Almond, D., J. Currie, and M. Herrmann. 2012. "From Infant to Mother: Early Disease Environment and Future Maternal Health." *Labour Economics* 19, no. 4: 475–83.
- Almond, D., L. Edlund, H. Li, and J. Zhang. 2010. "Long-Term Effects of Early-Life Development: Evidence from the 1959 to 1961 China Famine." In *The Economic Consequences of Demographic Change in East Asia*, ed. Takatoshi Ito and Andrew Rose, 321–45. NBER-EASE 19. Chicago: University of Chicago Press.
- Almond, D., L. Edlund, and M. Palme. 2009. "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden." *Quarterly Journal of Economics* 124, no. 4:1729–72.
- Azzoni, C. 2001. "Economic Growth and Regional Income Inequality in Brazil." *Annals of Regional Science* 35:133–52.
- Barker, D. 1992. *Fetal and Infant Origins of Adult Disease*. London: British Medical Journal.
- . 1997. "Maternal Nutrition, Fetal Nutrition and Diseases in Later Life." *Nutrition* 13:807–13.
- Barker, D., and L. Levy. 1994. *Mothers, Babies, and Disease in Later Life*. London: British Medical Journal.
- Barreca, A. 2010. "The Long-Term Impact of in Utero and Post-natal Exposure to Malaria." *Journal of Human Resources* 45, no. 4:865–92.
- Barros, F., C. Victora, and B. Horta. 2001. "Ethnicity and Infant Health in Southern Brazil: A Birth Cohort Study." *International Journal of Epidemiology* 30, no. 5: 1001–8.



- Bertrand, M., E. Dufflo, and S. Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119, no. 1: 249–75.
- Bleakley, H. 2010. "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics* 2:1–45.
- Bozzoli, C., A. Deaton, and C. Quintana-Domeque. 2009. "Adult Height and Childhood Disease." *Demography* 46, no. 4:647–69.
- Brabin, B. 1991. "The Risks and Severity of Malaria in Pregnant Women." Applied Field Research in Malaria Report no. 1, World Health Organization, Geneva.
- Breman, J. 2001. "The Ears of the Hippopotamus: Manifestations, Determinants, and Estimates of the Malaria Burden." *American Journal of Tropical Medicine and Hygiene* 64:1–11.
- Cameron, A., J. Gelbach, and D. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90, no. 3: 41427.
- Cameron, A. C., and D. L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50, no. 2:317–72.
- Catalano, R., C. Margerson Zilco, K. Saxton, and T. Bruckner. 2010. "Selection in Utero: A Biological Response to Mass Layoffs." *American Journal of Human Biology* 22:396–400.
- Causey, O., H. Penido, and L. Deane. 1943. "Observations on Malaria in the Presence and Absence of Anopheles Gambiae in an Experimental Area (Cumbe) Ceará, Brazil." *American Journal of Tropical Medicine and Hygiene* 23:59–71.
- Chang, S., B. Fleisher, S. Kim, and S. Liu. 2011. "Long-Term Effects of Early-Childhood Malaria Exposure on Education and Health: Evidence from Colonial Taiwan." IZA Discussion Paper no. 5526, Institute for the Study of Labor, Bonn.
- Chen, Y., and L. Zhou. 2007. "The Long-Term Health and Economic Consequences of the 1959–1961 Famine in China." *Journal of Health Economics* 26:659–81.
- Chiavegatto Filho, A., and I. Kawachi. 2013. "Are Sex-Selective Abortions a Characteristic of Every Poor Region? Evidence from Brazil." *International Journal of Public Health* 58, no. 3:395–400.
- Costa, D. 2012. "Scarring and Mortality Selection among Civil War POWs: A Long-Term Mortality, Morbidity and Socioeconomic Follow-Up." *Demography* 49, no. 4: 1185–1206.
- Crayen, D., and J. Baten. 2010. "Global Trends in Numeracy, 1820–1949, and Its Implications for Long-Term Growth." *Explorations in Economic History* 47, no. 1: 82–99.
- Cutler, D., W. Fung, M. Kremer, M. Singhal, and T. Vogl. 2010. "Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India." *American Economic Journal: Applied Economics* 2:72–94.
- Deane, L. 1988. "Malaria Studies and Control in Brazil." *American Journal of Tropical Medicine and Hygiene* 38, no. 2:223–30.
- Dehejia, R., and A. Lleras-Muney. 2004. "Booms, Busts, and Babies' Health." *Quarterly Journal of Economics* 119, no. 3:1019–1130.

- Dickinson, K., H. Randell, R. Kramer, and E. Shayo. 2012. "Socio-Economic Status and Malaria-Related Outcomes in Mvomero District, Tanzania." *Global Public Health* 7, no. 4:384–99.
- Domingues, P. 2010. "The Health Consequences of the Mozambican Civil War: An Anthropometric Approach." CES Working Papers, Centre d'Economie de la Sorbonne.
- Dulac, C. 2010. "Cerebral Malaria and Epilepsy." *Lancet Neurology* 9, no. 12:1144–45.
- Ewbank, D. 1981. *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. Washington, DC: National Academy Press.
- Galdo, J. 2010. "Long-Run Labor-Market Consequences of Civil War: Evidence from Shining Path in Peru." IZA Discussion Paper no. 5028, Institute for the Study of Labor, Bonn.
- Gallup, J., and J. Sachs. 2001. "The Economic Burden of Malaria." *American Journal of Tropical Medicine and Hygiene* 64, no. 1:85–96.
- Gradin, C. 2009. "Why Is Poverty So High among Afro-Brazilians? A Decomposition Analysis of the Racial Poverty Gap." *Journal of Development Studies* 45, no. 9:1426–52.
- Greene, W. 2012. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Harris, M., J. Consorte, J. Lang, and B. Byrne. 1993. "Who Are the Whites? Imposed Census Categories and the Racial Demography of Brazil." *Social Forces* 72, no. 2: 451–62.
- Holding, P., and P. Kitsao-Wekulo. 2004. "Describing the Burden of Malaria on Child Development: What Should We Be Measuring and How Should We Be Measuring It?" *American Journal of Tropical Medicine and Hygiene* 71:71–79.
- Holding, P., and R. Snow. 2001. "Impact of Plasmodium Falciparum Malaria on Performance and Learning: Review of the Evidence." *American Journal of Tropical Medicine and Hygiene* 64:68–75.
- Hong, S. 2007. "The Burden of Early Exposure to Malaria in the United States, 1850–1860: Malnutrition and Immune Disorders." *Journal of Economic History* 67:1001–35.
- Howson, C. 1996. *In Her Lifetime: Female Morbidity and Mortality in Sub-Saharan Africa*. Washington, DC: National Academies Press.
- Kelly, E. 2011. "The Scourge of Asian Flu: In Utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children." *Journal of Human Resources* 46, no. 4:669–94.
- Killeen, G., U. Fillinger, K. Ibrahim, L. Gouagna, and B. Knols. 2002. "Eradication of Anopheles Gambiae from Brazil: Lessons for Malaria Control in Africa?" *Lancet Infect Diseases* 2:618–27.
- Kudamatsu, M., T. Persson, and D. Strömberg. 2010. "Weather and Infant Mortality in Africa." CEPR Discussion Paper no. DP9222, Centre for Economic Policy Research, London.
- Kuecken, M., J. Thuilliez, and M. Valfort. 2015. "Does Malaria Control Impact Education? Evidence from Roll Back Malaria in Africa." HAL Archives. <https://hal.archives-ouvertes.fr/halshs-01099524/document>.

- León, G. 2010. "Civil Conflict and Human Capital Accumulation: The Long Term Effects of Political Violence in Peru." BREAD Working Paper no. 245, Bureau for Research and Economic Analysis of Development.
- Lindsay, S., J. Ansell, C. Selman, V. Cox, K. Hamilton, and G. Walraven. 2000. "Effect of Pregnancy on Exposure to Malaria Mosquitoes." *Lancet* 355:1972.
- Lindsay, S., and W. Martens. 1998. "Malaria in the African Highlands: Past, Present and Future." *Bulletin of the World Health Organization* 76:33–45.
- Lucas, A. 2010. "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka." *American Economic Journal: Applied Economics* 2:46–71.
- Maccini, S., and D. Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review* 99, no. 3: 1006–26.
- Macintyre, K. 2002. "Famine and the Female Mortality Advantage." In *Famine Demography: Perspectives from the Past and Present*, ed. T. Dyson and C. Ó Gráda, 240–60. Oxford: Oxford University Press.
- Marteletto, L. 2012. "Educational Inequality by Race in Brazil, 1982–2007: Structural Changes and Shifts in Racial Classification." *Demography* 49, no. 1:337–58.
- Martens, P., R. Kovats, S. Nijhof, P. de Vries, M. Livermore, D. Bradley, J. Cox, and A. McMichael. 1999. "Climate Change and Future Populations at Risk of Malaria." *Global Environmental Change* 9:S89–S107.
- McClelland, E., and J. Smith. 2011. "Gender Specific Differences in the Immune Response to Infection." *Archivum Immunologiae et Therapia Experimentalis* 59: 203–13.
- Mendis, K., B. Sina, P. Marchesini, and R. Carter. 2001. "The Neglected Burden of Plasmodium Vivax Malaria." *American Journal of Tropical Medicine and Hygiene* 64:97–106.
- Meng, X., and N. Qian. 2009. "The Long Term Consequences of Famine on Survivors: Evidence from a Unique Natural Experiment Using China's Great Famine." NBER Working Paper no. w14917, National Bureau of Economic Research, Cambridge, MA.
- Morlais, I., R. Girod, R. Hunt, F. Simard, and D. Fontenille. 2005. "Population Structure of Anopheles Arabiensis on La Reunion Island, Indian Ocean." *American Journal of Tropical Medicine and Hygiene* 73, no. 6:1077–82.
- Mu, R., and X. Zhang. 2011. "Why Does the Great Chinese Famine Affect the Male and Female Survivors Differently? Mortality Selection versus Son Preference." *Economics and Human Biology* 9:92–105.
- Oliveira-Ferreira, J., M. Lacerda, P. Brasil, J. Ladislau, P. Tauil, and C. Daniel-Ribeiro. 2010. "Malaria in Brazil: An Overview." *Malaria Journal* 9:115.
- Owusu-Agyei, S., D. Fryauff, D. Chandramohan, K. Koram, F. Binka, F. Nkrumah, G. Utz, and S. Hoffman. 2002. "Characteristics of Severe Anemia and Its Association with Malaria in Young Children of the Kassena-Nankana District of Northern Ghana." *American Journal of Tropical Medicine and Hygiene* 67, no. 4: 371–77.
- Parmakelis, A., M. Russello, A. Caccone, C. Marcondes, J. Costa, O. Forattini, M. Sallum, R. Wilkerson, and J. Powell. 2008. "Short Report: Historical Analysis of

- Near Disaster: *Anopheles Gambiae* in Brazil." *American Journal of Tropical Medicine and Hygiene* 78, no. 1:176–78.
- Rodrik, S., A. Subramanian, and F. Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9, no. 2:131–65.
- Ruckstuhl, K., G. Colijn, V. Amiot, and E. Vinish. 2010. "Mother's Occupation and Sex Ratio at Birth." *BMC Public Health* 10:269.
- Sachs, J., and P. Malaney. 2002. "The Economic and Social Burden of Malaria." *Nature* 415:680–85.
- Schellenberg, D., C. Menendez, E. Kahigwa, F. Font, C. Galindo, C. Acosta, J. Schellenberg, J. Aponte, J. Kimario, H. Urassa, H. Mshinda, M. Tanner, and P. Alonso. 1999. "African Children with Malaria in an Area of Intense *Plasmodium Falciparum* Transmission: Features on Admission to the Hospital and Risk Factors for Death." *American Journal of Tropical Medicine and Hygiene* 61, no. 3:431–38.
- Siraj, A., M. Santos-Vega, M. Bouma, D. Yadeta, D. Ruiz Carrascal, and M. Pascual. 2014. "Altitudinal Changes in Malaria Incidence in Highlands of Ethiopia and Colombia." *Science* 343, no. 6175:1154–58.
- Snow, R., C. Guerra, A. Noor, H. Myint, and S. Hay. 2005. "The Global Distribution of Clinical Episodes of *Plasmodium Falciparum* Malaria." *Nature* 434:214–17.
- Soper, F., and D. Wilson. 1943. *Anopheles Gambiae in Brazil, 1930–1940*. New York: Rockefeller Foundation.
- Steckel, H. R. 1991. "The Quality of Census Data for Historical Inquiry: A Research Agenda." *Social Science History* 15, no. 4:579–99.
- Störmer, C. 2011. "Sex Differences in the Consequences of Early-Life Exposure to Epidemiological Stress a Life-History Approach." *American Journal of Human Biology* 23:201–8.
- Strauss, J., and D. Thomas. 1995. "Human Resources: Empirical Modeling of Household and Family Decisions." In *Handbook of Development Economics*, vol. 3A, ed. T. N. Srinivasan, Jere Behrman, and Hollis B. Chenery, chap. 34. Amsterdam: Elsevier.
- Trape, J., and C. Rodgier. 1996. "Combating Malaria Morbidity and Mortality by Reducing Transmission." *Parasitology Today* 12, no. 6:236–40.
- Trivers, R., and D. Willard. 1973. "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring." *Science* 179, no. 4068:90–92.
- Valente, C. 2015. "Civil Conflict, Gender-Specific Fetal Loss, and Selection: A New Test of the Trivers-Willard Hypothesis." *Journal of Health Economics* 39:31–50.
- van Lieshout, M., R. Kovats, M. Livermore, and P. Martens. 2004. "Climate Change and Malaria: Analysis of the SRES Climate and Socio-Economic Scenarios." *Global Environmental Change* 14:87–89.
- Venkataramani, A. 2010. "Early Life Exposure to Malaria and Cognition and Skills in Adulthood: Evidence from Mexico." *Journal of Health Economics* 31, no. 5:767–80.
- Vlassoff, C., and E. Bonilla. 1994. "Gender Related Differences in the Impact of Tropical Diseases on Women: What Do We Know?" *Journal of Biosocial Science* 26:37–53.

- Waldron, I. 1983. "Sex Differences in Human Mortality: The Role of Genetic Factors." *Social Science and Medicine* 17, no. 6:321–33.
- Webb, M. 2013. "Reworking Wild Bootstrap Based Inference for Clustered Errors." Economics Department Working Paper no. 1315, Queen's University.
- Wells, J. 2000. "Natural Selection and Sex Differences in Morbidity and Mortality in Early Life." *Journal of Theoretical Biology* 202:65–76.
- WHO (World Health Organization). 1998. "Malaria Epidemics Detection and Control Forecasting and Prevention." Report WHO/MAL/98.1084.
- . 2000. "Severe Falciparum Malaria." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 94:1–13.
- . 2008. *World Malaria Report, 2008*. Geneva: WHO.
- . 2013. *World Malaria Report, 2013*. Geneva: WHO.
- Wood, C., J. de Carvalho, and C. Horta. 2010. "The Color of Child Mortality in Brazil, 1950–2000: Social Progress and Persistent Racial Inequality." *Latin American Research Review* 45, no. 2:114–39.
- Worrall, E., S. Basu, and K. Hanson. 2005. "Is Malaria a Disease of Poverty? A Review of the Literature." *Tropical Medicine and International Health* 10, no. 10: 1047–59.