Shortened Lifespan: A Legacy of Exposure to Malaria Risk in Early Life

Sok Chul Hong *
Seoul National University

Abstract

This paper studies historical experience to measure a long-term potential benefit of malaria eradication. Many Americans in the first half of the 19th century were exposed to high risk of malarial fevers because medical knowledge for effective prediction and prevention was inadequate. Using the sample of Union Army veterans born during the period and their lifetime records, this study examines that the exposure to high risk of malaria at birth or in early life substantially shortened their lifespan. The legacy is estimated robust with controlling for lifetime socioeconomic and health conditions, fixed effects and selection. The negative impact on lifespan is found more substantially at veterans' younger ages, and more frequently through certain types of causes such as respiratory and digestive diseases.

Keywords: Malaria, Lifespan, Nineteenth-Century America, Union Army Veterans
JEL Classification: I15, J10, N31,

* Associate Professor, Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826. Email: sokchul.hong@snu.ac.kr

Research reported in this article was supported by NIH grant number P01 AG10120.

1. Introduction

Malaria is the world's important parasitic infectious disease, which is transmitted by mosquitoes. According to WHO's World Malaria Report, an estimated 216 million cases of malaria occurred worldwide and 445,000 people died in 2016, mostly children in the African region; about 3.3 billion people still live in 97 countries and territories at the risk of malaria transmission in 2016 (WHO 2017). The eradication of malaria is one of the global priorities not only because of its high mortality and prevalence rates, but also because malaria is known to cause adverse health and economic outcomes such as low birthweight, severe anemia, brain damage, kidney failure, and loss of school and working days (WHO 2017). Economic studies have emphasized that malaria infection especially in early life can deteriorate cognitive ability in growing years, lower labor productivity and income in adulthood, and accelerate the onset of chronic conditions in old ages (Barreca 2010; Bleakley 2010; Chang et al. 2011; Cutler et al. 2010; Hong 2007, 2011, 2013; Lucas 2010). Those studies strongly suggest that the benefit of malaria eradication is tremendous in the long-term perspectives.

This study examines the effect of malaria in a longer-term perspective by investigating an unexplored question: can the exposure to malaria risk in early life shorten lifespan among infected survivors? Economic studies above suggest that lifespan can be affected through indirect mechanisms in which malaria lowers socioeconomic status and lifetime health. In addition, a recent biological study on wild birds reveals that malaria infections significantly shorten the lifespan of infected birds through a biological mechanism in which malaria accelerates telomere degradation, which is closely associated with lifespan and aging process (Asghar et al. 2015).

The effect of malaria on human beings' lifespan has been rarely studied. Studies on lifespan require longitudinal data containing lifetime information on exposure to malaria, lifespan and various confounding factors for identification. However, it is quite difficult to find modern data that satisfy such

conditions. This study overcomes the limitation by utilizing the historical records on Union Army veterans who attended the American Civil War in 1861-65 and their early-life experience of malaria in the mid-19th century. Malarial fever had been prevalent across most areas of the United States throughout the period as much as today's African countries, and its pathogen was not clearly revealed until 1898. Thus, this historical experience provides a framework of natural experiment. In addition, the Union Army dataset contains veterans' lifetime variables for examining the effect on lifespan and for controlling for confounding factors.

This study conducts various analyses. First, I limit the Union Army sample to veterans who survived up to 1900 (or age 60 on average) to fix a potential bias due to selection, and measure how severely each veteran was exposed to malarial fevers at the county of birth. Then, I estimate that their lifespan was significantly affected by the risk at birth using state-of-birth fixed effects. Those born in counties with a one-standard-deviation (1-SD) higher risk of malaria lived a 0.84-year shorter than those born in average counties. The result is robust with controlling for wartime experience, chronic health conditions, and endogeneity problems.

Second, I show that the effect on lifespan is substantial even though the variables of socioeconomic status at various points of lifetime are added to the regressions. In addition, the effect of exposure to malaria risk is significantly estimated only for the county of birth, but insignificant for the exposure at ages 10, 20 and 60. Third, I extend the sample by adding veterans who died before 1900 on the regressions. Then, the negative impact of malaria on lifespan is estimated more substantial at younger ages. For example, the marginal effect of 1-SD malaria risk is estimated a 2.3 year when all the war survivors are included. This suggests that veterans born in malarial counties died very early at young ages. Finally, I estimate that veterans born in high-malaria counties more likely died from certain types of causes such as respiratory and digestive diseases and neoplasm.

The estimates in this paper are robust and consistent with controlling for other key controls like socioeconomic and health variables measured over a course of lifetime. This may suggest that the negative effect on lifespan occurred not only through accumulative poor socioeconomic and chronic conditions, but also through biological mechanism as revealed for wild birds by Asghar et al. (2015). But how malaria in early life can shorten human beings' lifespan requires further rigorous investigations. Rather than its mechanism, this study will be significant in that it quantitatively measures the effect of malaria on lifespan for human beings. The findings in this study will provide new insight for measuring the potential benefit of malaria eradication and for understanding the significance of early-life exposure to infections in the past.

2. Background and Related Literature

Economists have provided empirical evidence showing that early-life exposure to malaria can have the long-term impact on economic and health-related outcomes over a course of lifetime. To identify its causality, many studies have utilized countries which experienced malaria eradication by the mid-twentieth century, and examined whether cohort born after the eradication significantly had better outcomes than did the preceding cohorts. Using panel data consisting of year-of-birth by state-of-birth cohorts from multiple historical censuses in the United States, Brazil, Colombia and Mexico, Bleakley (2010) estimated that cohorts born after eradication had higher income as adults than did the preceding cohorts. Cutler et al. (2010) found that malaria eradication improved educational attainment in India. Similar implication was found for Paraguay and Sri Lanka by Lucas (2010), and for Mexico by Venkataramani (2012). In the aspect of health outcomes, Hong (2013) studied that US cohorts exposed to an intensive level of the anti-malaria campaign, which began in 1921, had a lower probability of having work disability in old age. Using the longitudinal health records of Union Army sample generally born

around 1840, Hong (2013) also estimated that exposure to a malarial environment in their early life substantially increased the likelihood of having certain chronic diseases and not working around age 60.

Longevity is known to be determined by various non-genetic factors that an individual experiences over the lifespan including socioeconomic status, health-related behaviors and health conditions (Christensen and Vaupel 1996). Thus, those economic studies above strongly imply that poor human capital and adverse health consequences driven by and accumulated after early exposure to high risk of malaria can shorten lifespan. Many studies have supported this hypothesis in general context. Using the data of Health and Retirement Study, Montez and Hayward (2014) showed that early-life disadvantages in terms of socioeconomic status and health both shortened lives and increased the years lived with functional impairment. In the biological aspect, Crimmins and Finch (2006) showed that the decline in old-age mortality was prompted by the reduced burden of infections and inflammation especially at younger ages. Some studies suggest an intergenerational transmission of socioeconomic status, health and even longevity through such indirect channels (Case et al. 2002). Accordingly, the link between malaria in early life and shortened lifespan looks very probable.

The link also has been backed up by a recent biological study of Asghar et al. (2015). They investigated a wild population of great reed warblers which nest in Sweden and overwinter in Africa. Interestingly, those migrant birds can pick up various species of malaria parasites while overwintering in the tropics. So after initial acute malaria, they are asymptomatically infected for life with so-called chronic malaria. Using 25-year data on great reed warblers, the research team discovered that chronic malaria infection shortened the first generation's lifespan and reduced the lifetime number and quality of their offspring. Asghar et al. (2015) also found that infected birds had significantly shorter telomeres (the

protective caps on the ends of chromosomes), which is known to be associated with short lifespan and aging process (Heidinger et al. 2012).[1]

The findings of the biological study above have important implications. If the link between malaria in early life and lifespan is mostly explained by indirect channels through poor human capital and impaired health accumulated over a lifetime, social programs that support affected individuals can be useful to alleviate such negative long-term effects. However, if the link results from phenotypic change--- telomere degradation, such compensating policies would be less effective or limited in reducing the hidden cost. In the case, malaria eradication should be considered more significant.

Little is known about the effect of malaria on human being's lifespan among survivors. Althogh only several studies regarded the subject, their scope is very limited. Some relevant studies seem very limited. Bawah and Binka (2007) estimated that, if malaria is eradicated from a hyper-endemic area of Northern Ghana, life expectancy at birth would increase by more than six years. But this study considered only life losses directly from malaria infections. Using the data of elderly population in Costa Rica in 2004, Brenes-Camacho and Palloni (2011) found that survivors from malaria infections in early life more likely died due to stroke than otherwise. But the association with lifespan was not examined. Lifespan is a longer-term outcome that reflects various aspects of lifetime experiences and quality of life. Thus, the current study on how lifespan is affected by early-life exposure to malaria would provide new insight for measuring the potential benefit of malaria eradication and similar infectious diseases.

3. Malaria in Mid-19th Century America, Union Army Sample, and Variables

3.1 Historical Background

[1] The study discovered that the offspring of infected birds also had shortened telomeres. This suggests that chronic malaria infection can impair phenotypic quality and Darwinian fitness.

It is known that malaria was introduced into North America by early European migrants in the 16th century and African slaves in the 17th century. Since then, malarial fevers had been very prevalent across most areas of the United States throughout the mid-19th century. Although some scientists suspected that malarial fevers might be associated with warm and wet weather, its transmission via mosquitoes was not revealed until 1898. This implies that those in the areas with climate suitable for mosquitoes were exposed to high risk of malarial fevers without effective prevention and treatment (Humphreys 2001).

The term of malaria was not used in the 19th century. Malarial fevers were distinguished from fevers due to other infectious diseases by their pattern of fevers. Remittent and intermittent fevers were classified into malarial fevers. According to the 1850 population census, the 1850 mortality rate of malarial fevers per 100,000 was 68 for all the states and 111 for 7 southern states such as KY, TN, AL, MS, AR and TX (U.S. Census Bureau 1855). The average malarial mortality rate among African countries in 2016 is 43 per 100,000 (WHO 2017). Therefore, mid-19th century America provides a useful historical frame to examine the effect of exposure to malaria risk.

3.2 Union Army Sample, Key Variables and Selection of Sample

This study uses the sample of Union Army cohort that consists of nearly 40,000 white soldiers from 331 companies who attended the American Civil War (1861-1865). The sample has been collected for the *Early Indicators* project. Native soldiers were born around 1840 on average. Thus, many of them were exposed to high risk of malarial fevers in early life. Each soldier in the sample has been linked to various historical records including the Complied Military Service Records (CMSR), Pension Records (PEN), Carded Medical Records (CMR), Surgeons' Certificates (SCRTS), and US Federal Censuses in 1850-1940 (CEN). The combination of the records provides each soldier's lifetime information on birth

(from CMSR), early-life conditions (from CEN), wartime experience (from CMSR), late-life conditions (from CEN), health conditions during the War (from CMR) and after the War (from SCRTS), death (from PEN) and so on.[2]

The key control variable in this study is the level of the exposure to malarial fevers in early life. Three issues on the variable need to be addressed. First, the dataset above does not provide information on whether observations contracted malarial fevers in early life or not. Even if we know the dummy of infection in some points in early life, it would not measure how soldiers frequently were infected with malaria because former infection generally does not give immunity. Instead, I estimate the annual probability of contracting malarial fevers in the county of birth, in 1850 (about age 10 on average), in 1860 (about age 20) and in 1900 (about age 60) in order to measure how much soldiers were exposed to malaria in early life or late life. The county-level indexes of malaria risk are estimated on the basis of county-level temperature, rainfall and elevation. In particularly, I estimated the index for birth county using average weather variables for 1825-1859. This is for utilizing more reliable weather variables. So in the baseline estimation, I do not consider variations between years of birth, but variations between places of birth. In other words, the index can be said to measure the level of chronic malaria infections. Its estimation method was first suggested in Hong (2007), and more details are provided in Appendix A.

Second, I will use the term of 'early life' to denote the years of age from birth up to age 10. Considering existing modern studies that emphasize the significance of malaria infection during mother's pregnancy, the in-utero period is also contained into the range of early life in this paper. The UA dataset does not provide residential information on every movement in early life. Only the counties at birth, in 1850 (age 10 on average) and in 1860 (age 20 on average) are available for the period prior to the Civil War. As reported in next section, the effect of malaria risk at birth on lifespan is estimated significant, and

[2] More detailed information on the Union Army sample and the Early Indicators project is available at uadata.org.

the significance is not estimated for ages 10 and 20. But the UA dataset does not allow to test the significance for other ages between birth and age 10.

Third, as discussed above, this study emphasizes the exposure to malaria risk in the county of birth. Thus, all the analyses below use only native-born veterans whose birth counties are known. The malaria risk in 1850 and 1860 is estimated for veterans' residential counties reported in the 1850 and 1860 census manuscripts. Thus, when the estimated malaria risk in 1850, 1860 or 1900 is contained in regression analyses, the sample will be confined to those who are linked to each year's census manuscripts.

The outcome variable in this study is lifespan, i.e., age at death. The variable has two issues. First, about 15% of UA veterans died during the Civil War. Although many died from external causes like injuries during battles, it seems less clear whether deaths from other causes were affected by early-life exposure to malaria risk. Thus, I exclude those who died during the War from the baseline estimation in Section 4.1. Instead, they will be considered in the analysis of robustness check in Section 6 and discuss related issues.

Second, the availability of lifespan variable is related with the change of Civil War Pension law. The early pensions for Civil War veterans was given to only those who could prove time spent in the military and had a disability incurred while in service. Accordingly, only a limited number of veterans could enroll in the pension system. The system significantly changed in 1890, when the Congress passed the Dependent and Disability Pension Act. The new act provided pensions for all veterans who had serviced at least 90 days in the Union military and were honorably discharged, regardless of war-related disability. After 1890, the number of veterans who enrolled in the pensions system substantially increased.[3] The problem is that the information on veterans' death year is available only from their pension

[3] About 39% of war survivors (12,974 out of 33,661) applied for the pension until 1889. Then, 4,045 (12% of war survivors) applied for the pension in 1890, which is the year when the highest number of applications is observed.

records when they survived the Civil War. This means that if veterans did not have war-related disabilities and died before 1890, they could not enroll in the pension system and their birth year is not available. If deaths prior to 1890 were affected by early-life malaria, but such veterans are excluded from analyses, the effect on lifespan can be underestimated. To correct this potential bias, the baseline estimation in Section 4.1 uses only native-born veterans who survived up to 1900 and so could apply for the pension.[4] I release this constraint in Section 6 and discuss related issues.

4. Identifying the Effect of Malaria Risk at the County of Birth

4.1 Correlation between Lifespan and Malaria Risk at Birth

As the first step of baseline analysis, I examine the relationship between lifespan and malaria risk measured for the county of birth. Following the discussion above, only 11,006 native-born veterans whose birth county and age at death are known and who survived up to 1900 are selected for the baseline analysis. I clustered the selected sample into 917 cohorts born in the same county. Figure 1 shows the scatter plot between each cohort's average lifespan and average malaria risk at birth. The dashed line is added to see the linear relationship between two variables in the pooled data. It suggests that lifespan was not related with malaria risk at birth.

[Figure 1 Here]

However, the pooled-data relationship seems misleading because this does not capture the disparities of local characteristics across states or regions. I added the linear fitting (solid) lines estimated

[4] I use the year of 1900 rather than 1890. This is because many veterans newly applied for the pension throughout the 1890s. So most veterans are thought to have applied for the pension by 1900. In fact, only 4% of veterans newly applied for the pension after 1900.

for each state-of-birth sample in the figure. For most states, negative correlation is observed. This implies that state-of-birth fixed effects need to be contained in regression analyses below.

4.2 Baseline Estimation

I conduct the baseline estimation using the following equation:

$$L_{ics} = \alpha + \beta M_c + \delta_s + X_c\Gamma + \sum_t X_{i,t}\Pi_t + \varepsilon_{ics} \quad (1)$$

In the equation, I assume that the lifespan (L_{ics}) of a veteran i born in state s's county c mainly depends on his birth county's malaria risk (M_c). I normalize the malaria ecology index for easy interpretation of the estimated results, but this convert does not make any changes to the main findings and their significance compared with those done with the index not normalized. Following the lesson from Figure 1, I add the state-of-birth fixed effects. This will not only produce within-state estimators, but also capture time-invariant state-of-birth factors.

I also consider various confounding factors that could affect veterans' lifespan. First, I control for early-life local conditions measured for the county of birth (X_c). They include crude death rate, mortality rate due to 9 major infectious diseases, infant mortality rate, population density, fertility rate, school enrollment rate, average farm value, ratio of slave populations, and dummy of whether counties were connected by railroad. I obtained these county variables from the 1850 census publications because many variables are not available prior to 1850. Second, I also control for veterans' personal characteristics ($X_{i,t}$) such as year of birth, variables that measure wartime experiences, and dummies that indicate the onset of chronic diseases by 1899. The variables of wartime experiences include year of enlistment, dummies that indicate whether veterans had diseases or injuries, initial military rank, dummy of prisoner-of-war

experience, and wartime mortality rate within company. The onset of chronic diseases is measured for 18 types of diseases, which are recorded in surgeons' certificates. On the other hand, other kinds of personal socioeconomic variables are available from the census records. They will be considered in the next subsection; the baseline regressions here do not utilize census links.

Table 1 reports the estimated coefficients of the key control variable, malaria ecology risk at the county of birth, and their standard errors clustered on the county of birth. Column (1) controls for only malaria risk and birth year. As expected from the dashed line in Figure 1, the coefficient of column (1) estimated without state-of-birth FE is positive and statistically insignificant. However, the use of state-of-birth FE in columns (2)-(5) strongly suggests that veterans born at more malarial counties had shorter lifespan than otherwise. The result is robust even though other controls are added. In terms of the coefficient in column (5), being born in a county with one-standard-deviation higher malaria risk may reduce lifespan by 0.84 year. In the dataset, the most and least malarial counties had a gap of about 4 standard deviations in malaria index. This implies that more than 3-year gap in lifespan can be caused by the gap.

[Table 1 Here]

4.3 Endogeneity and IV Estimation

The baseline estimation above assumes that malaria risk at birth was not endogenously determined. Because the pathogen of malaria was not revealed in the mid-19th century, it is reasonable to assume that people (or veterans' parents) had selected counties without adequate information on the local risk of malarial fevers. In addition, the malaria ecology index is estimated on the basis of exogenous environmental factors; the variable is a kind of instrumental variable for actual malaria risk. However,

some historical studies suggest that some people in the mid-19th century suspected the relevance of hot/wet weather to malaria fevers from their experience, even though they never knew the role of mosquitoes. This may overestimate the effect of malaria on lifespan if families with lower socioeconomic status migrated to more malarial counties. In addition, omitted variables correlated with malaria risk at the county of birth can cause another type of endogeneity problem.

I examine the existence of selection and endogeneity by conducting two alternative estimation models. First, instead of state-of-birth FEs, I use county-of-birth FEs in column (6) of Table 1. In order to utilize within-county variation, I re-estimate malaria risk for each individual year of birth by using annual temperature and accumulated precipitation rather than 1825-1859 average.[5] This specification is very strict because it uses within-county variations. Nevertheless, the result in column (6) is comparable with the baseline results.

Second, in columns (7) and (8), I try to correct the potential selection problem by using the ratio of adult populations in veterans' county of birth who came from the same U.S. states or countries as those of veterans' fathers for an instrumental variable. Many southern areas in early-19th century America were unexplored, hot and wet, and so at the high risk of malarial fevers. However, better agricultural and economic conditions attracted immigrants (Hong 2011). People needed information on places to immigrate into, and they relied on their friends or ethnic groups who already migrated into the places (Costa and Kahn 2006). The instrumental variable assumes that veterans' fathers migrated into counties with populations from the same states or countries as fathers' and that those counties were malaria. Because detailed information on county populations' birth places is available from the 1850 population census, I calculate the instrumental variable using the 1850 IPUMS data.

[5] As discussed in Appendix A, weather information in a specific year prior to 1860 is inadequate and less reliable because the number of weather stations was small.

The result of the first stage in column (7) well supports the above assumption for the instrumental variable, and shows that it is statistically valid. In the IV estimation of column (8), I obtain a similar size of coefficient for malaria-risk variable to that of baseline estimation. Although it is statistically insignificant, a test rejects the hypothesis that counties (in the respect of malaria risk) were selected endogenously. This may be affected by that sample size decreased in the process of calculating the instrumental variable. In column (9), I conduct the baseline estimation (baseline 2 in Table 1) using 6,695 veterans used for IV estimation, and its result is similar with that of column (5). In summary, although alternative estimations in columns (6)-(9) may be imperfect to correct the potential problems of selection and omitted variables, the results above suggest that the problems are minor or do not damage the result of baseline estimations.

5. Significance of the Exposure to Malaria Risk from Birth up to Age 10

In this section, I utilize the 1900, 1860, and 1850 census records to which veterans are linked. The links are useful to see whether the effect of malaria risk is robust even though veterans' socioeconomic status is controlled for. In addition, the information on residence found in the census records will be utilized to test the significance of exposure to malaria risk at various ages. Throughout this section, other specifications---control variables and fixed-effects model---are the same as those used in column (5) of Table 1.

Columns (1)-(3) in Table 2 use 7,592 veterans who survived up to 1900 and were also found in the 1900 census manuscripts. Column (1) use the same regression specification as that of column (5) in Table 1, and column (2) additionally controls for veterans' socioeconomic status around 1900 (or age 60 on average) including marital status, literacy, and home ownership. Column (2) suggests that the effect of

malaria risk in early life is robust even though socioeconomic conditions in old ages are controlled for. In column (3), I add the malaria ecology index at the 1900 residential county estimated with the 1900s average weather variables. The result shows that the exposure to malaria risk in early life is still robust, and the exposure around age 60 did not affect veterans' lifespan.[6]

[Table 2 Here]

In columns (4)-(7), I use 3,894 veterans who survived up to 1900 and were found in both 1900 and 1860 census manuscripts. A high probability of being found in multiple censuses implies that the veteran has more records in the dataset to use for census linkages probably because he might be less affected by malaria and might live longer. This possibility of positive selection may explain why the coefficient in column (4) is estimated somewhat higher than that of the baseline estimation in Table 1. Column (5) shows that the effect is robust even though household head's wealth and family size around 1860 (or age 20 on average) are added as control variables. Columns (6) and (7) also control for the malaria ecology index at the 1860 residential county estimated with the 1860s average weather variables. Only the exposure to malaria risk at the county of birth is estimated substantial and statistically significant.

I add one more constraint in choosing sample for columns (8) and (9) by using the 1850 census links. Thus, the size of sample decreased to 2,621. The effect of malaria risk at the county of birth is still robust even though I control for household head's real estate wealth and family size and malaria ecology

[6] The 1900 malaria ecology index is estimated on the basis of the relationship between actual malarial prevalence rate in military camps in the early and mid-19th century and their environmental factors in the same period. Therefore, the 1900 ecology index may not reflect actual malaria risk around 1900 well. In addition, it is known that malarial fevers began to be retreated from the mid-west region in the 1880s because of large-scale farm land reclamation. Accordingly, the risk of malarial fevers around 1900 was low, and veterans' lifespan was less affected by the fevers.

index in 1850 (or age 10 on average). Many veterans lived in the county of birth in 1850; the ecology index at the county of birth is highly correlated with that in 1850. Nevertheless, only the coefficient of malaria ecology index at the county of birth is estimated significant.

Identifying the effect of exposure to malaria risk between birth and age 10 is not possible because the data set does not provide residential information before 1850. Similarly, it is impossible to identify the effect of in-utero exposure due to the limitation of dataset. Thus, the significant coefficient of malaria ecology index at the county of birth in this paper may reflect the effect at different ages from in-utero period up to age 10.

6. Larger Effects at Younger Ages

So far, I have limited the sample to those who survived up to 1900 to fix the potential bias from that many veterans died without the records on death year before the general pension law passed in 1890. However, this can underestimate the effect of early-life exposure to malaria risk on lifespan if those who were exposed to a higher risk of malaria around birth died earlier before 1900. To test this possibility, I choose veterans who survived up to every 5 year t from 1866 (right after the Civil War) to 1925, and respectively, conduct the baseline estimation (baseline 2 in Table 1) for each sample.

In the left panel of Figure 2, I plot the coefficients of malaria ecology index at the county of birth estimated from the 13 separate regressions (denoted by solid line) and their 90% confidence intervals (denoted by dashed lines). The x axis in the figure denotes the year t used for choosing veterans.

[Figure 2 Here]

The result in the figure strongly suggests that veterans from malarial counties died early before 1900. The marginal effect of one-standard-deviation malaria risk was estimated a 0.84 year among those who survived up to 1900 in Section 4.2. However, the marginal effects are estimated a 1.17 year among survivors by 1890, a 1.75 year among survivors by 1880, a 2.07 year among survivors by 1870, and a 2.30 year among survivors by 1866; all the coefficient is statistically significant. The marginal effect declines as veterans who survived by age older than 60 are selected, and becomes statistically insignificant.

I also test the above issue using the Cox proportional hazard estimations. I use the same control variables and fixed effects as in the baseline estimations; various samples of veterans are employed as done in Figure 2 (a). Each regression's coefficient and its 90% confidence intervals are graphed in Figure 2 (b). As the coefficient of malaria ecology index in the hazard regression is higher than 1, it is interpreted that veterans from malarial counties died more quickly than otherwise. Therefore, the hazard model provides the same implication as that of Figure 2 (a).

Both estimation results in Figure 2 strongly suggests that the negative health impact of early-life exposure to malaria risk is stronger when veterans were younger. Only a small number of veterans from malarial counties are thought to have survived up to old ages, and so the negative impact on lifespan is estimated smaller among the old-age survivors.

7. Frequent Cause of Death among Veterans from High-Malarial Region

Finally, I examine whether veterans' cause of death was influenced by early-life exposure to malaria risk. Using UA veteran sample, Hong (2013) found that those born in more malarial counties had more chronic conditions at old ages than otherwise. Particularly, the likelihood of having digestive-system related health problems was estimated higher among those from more malarial counties. This suggests

that malaria may impair the function of body and specific organs, and so shortened lifespan estimated above could be associated with those health impairment and chronic conditions.

For this analysis, I utilized veterans' death certificates that contain the information on cause of death. The *Early Indicators* project coded each cause of death recorded in the certificates using WHO's ICD-10 code. I use the dummy variable indicating whether the veteran died from specific cause or not as dependent variables. Other specification is the same with those in the baseline estimations used in Table 1 (column 4 or 5).

Table 3 reports the estimated coefficients and standard errors of malaria ecology index measured at the county of birth. I conducted the linear probability regressions for nine types of leading causes of death among veterans; some causes with low frequencies such as mental, musculoskeletal or sensor-related diseases were excluded.

[Table 3 Here]

Two findings are worthy to note. First, veterans born in malarial counties more likely died from several specific causes than otherwise. In terms of the coefficients in column (7), being born in a county with one-standard-deviation higher malaria risk from counties with average index significantly increased the probability of dying from respiratory diseases by 5.5 % points, digestive diseases by 2.6 % points, and neoplasm by 2.5% points. The magnitude of each coefficient is substantial accounting for about half of sample mean, respectively. The findings are very consistent across the sample used and speciation, especially regardless of controlling for chronic conditions by 1899 between columns (6) and (7). Second,

Table 4 shows that veterans born in high-malaria counties less likely died from senility. This implies that they more frequently suffered from chronic conditions and diseases at the end of their lifetime.

The result above is quite consistent with Hong's (2013) findings that those born in malarial counties had more chances of having gastrointestinal diseases, rectum, and hemorrhoids, which are classified as digestive diseases classified in this study. It is also intriguing to know that biological studies have revealed that telomere degradation may be risk factors for respiratory and digestive diseases and cancers (Calado 2014; Zhu et al. 2016). The length of telomere was pointed out as a key mechanism through which chronic malaria can shorten lifespan among great reed warblers in Asghar et al. (2015). Accordingly, it is suggestive that malaria in early life may lead to shortened lifespan among human beings via certain biological channels. But its investigation is beyond of this current study's scope.

8. Concluding Remarks

Although early-life exposure to malaria has been well studied to affect human capital accumulation and lifetime health, little is known about how malaria experienced in early life can affect lifespan among survivors. Using the longitudinal records of Union Army veterans who were born around 1840, when malarial fever was very prevalent, I estimated that those born in high-malaria-risk counties more likely died more quickly than otherwise. I also found that the effect on lifespan occurred more substantially at their young ages, and that deaths from respiratory and digestive diseases were more frequently found among those from more malarial counties. Although this study does not reveal exact mechanism in which malaria in early life can shorten remaining lifespan, the legacy of malaria can be explained by lifetime interactions among biological, environmental and socioeconomic factors and through direct and indirect channels.

The findings in this study have several implications. First, from the aspect of global health, they well support that the potential benefit of malaria eradication would be larger than discussed so far. In particular, recent biological studies suggest that telomere degradation due to chronic malaria may shorten lifespan. They also show that this channel can work over generations by reducing offspring's' lifespan. In the case, later public and private supports for compensating impaired health can be less effective. This makes malaria eradication more significant. Second, the findings in this study provide empirical evidence accounting for why people in the past died earlier or had short lifespan. High infant mortality rate mostly due to poor sanitation and infections is the most direct explanation. But even if people survived from infections, its impact is long-lasting affecting remaining lifespan.

Appendix A: Estimation of Malaria Ecology Index

To estimate the county-level malaria ecology index, I consider various risk factors for malarial fever following the estimation method from Hong (2007): temperature, precipitation during the months when mean temperatures were higher than 59 degrees Fahrenheit, standard deviation of elevation (as a proxy of wetland or swamps), the extent to which land was improved for agriculture and whether or not the region is adjacent to the ocean. The risk estimation consists of two steps. The first is to calculate the correlation between the malaria incidence rate among the 143 U.S. Army forts in 1829-1874 and the risk factors around the forts. The second step is to estimate the county-level malaria risk by plugging county-level risk factors into the regression result of the first step. Thus, the estimated index can be considered as representing the annual probability that people in the given county would contract malarial fevers.

Figure A1 shows the malaria ecology index estimated with average weather variables for 1825 to 1859, which has been used as the key control variable in this study. The maximum and minimum of the estimated index are 0.5623 and 0.0032, respectively. Its mean and standard error are 0.2398 and 0.1155,

respectively. For column (6) in Table 1, I estimated the indices for each county and year for the period from 1825 to 1850 on the basis of each year's weather variable. For the malaria ecology index in 1850, 1860 and 1900 used in Table 2, I used the average weather variables for 1850 to 1859, 1860 to 1869 and 1900 to 1909, respectively. The historical monthly weather variables were obtained from National Climatic Data Center and the US Historical Climatology Network. The Kriging estimation method, which is a geospatial technique, was applied to estimate weather variables across the entire US area. Detailed data sources and related-issues are discussed in appendix of Hong (2007).

[Figure A1 Here]

Appendix B: Control Variables

Table A1 shows the description and source of the control variables used in the baseline estimations.

[Insert Table A1 Here]

References

Asghar, M., D. Hasselquist, B. Hansson, P. Zehtindjiev, H. Westerdahl, and S. Bensch. 2015. "Hidden Cost of Infection: Chronic Malaria Accelerates Telomere Degradation and Senescence in Wild Birds." *Science* no.347(6220): 436-438.

Barreca, Alan. 2010. "The Long-Term Economic Impact of in Utero and Postnatal Exposure to Malaria." *Journal of Human Resources* no. 45 (4): 865-892.

Bawah, Ayaga A., and Fred N. Binka. 2007. "How Many Years of Life Could Be Saved If Malaria Were Eliminated from A Hyperendemic Area of Northern Ghana?" *American Journal of Tropical Medicine and Hygiene* no.77(Suppl 6)145-152.

Bleakley, Hoyt. 2010. “Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure.” *American Economic Journal: Applied Economics* no. 2 (2): 1-45.

Brenes-Camacho, Gilbert, and Alberto Palloni. 2011. “Malaria Survivors during Early Life, Health at Old Age, and Stroke Mortality in Costa Rica.” Paper presented at the Annual Meeting of the PAA, Washington, DC.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. “Economic Status and Health in Childhood: The Origins of the Gradient.” *American Economic Review* no.92(5): 1308-1334.

Calado, Rodrigo T. 2014. “Telomeres in Lung Diseases.” *Progress in Molecular Biology and Translational Science* no.125: 173-183.

Chang, S., B. Fleisher, S. Kim, and S. Liu. 2014. “Long-Term Effects of Early Childhood Malaria Exposure on Education and Health: Evidence from Colonial Taiwan.” *Economic Development and Cultural Change* no. 62(3): 519-536.

Christensen, K., and J.W. Vaupel. 1996. “Determinants of Longevity: Genetic, Environmenal and Medical Factors.” *Journal of Internal Medicin*e no. 240: 333-341.

Costa, Dora, and Matthew Kahn. 2006. “Forging a New Identity: The Costs and Benefits of Diversity in Civil War Combat Units for Black Slaves and Freemen.” *Journal of Economic History* no. 66(4): 936-962.

Crimmins, Eileen M., and Caleb E. Finch. 2006. “Infection, Inflammation, Height, and Longevity.” *PNAS* no.103(2): 498-503.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl. 2010. "Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India." *American Economic Journal: Applied Economics* no. 2 (2): 72-94.

Heidinger, Britt J., Jonathan D. Blount, Winnie Boner, Kate Griffiths, Neil B. Metcalfe, and Pat Monaghan. 2012. "Telomere Length in Early Life Predicts Lifespan." *PNAS* no.109(5): 1743-1748.

Hong, Sok Chul. 2007. "The Burden of Early Exposure to Malaria in the United States, 1850-1860: Malnutrition and Immune Disorders." *Journal of Economic History* no. 67 (4): 1001-1035.

———. 2011. "Malaria and Economic Productivity: A Longitudinal Analysis of the American Case." *Journal of Economic History* no. 71 (3): 654-671.

———. 2013. "Malaria: An Early Indicator of Later Disease and Work Level." *Journal of Health Economics no*. 32 (3): 612-632.

Humphreys, Margaret. 2001. *Malaria: Poverty, Race, and Public Health in the United States*. Baltimore: The Johns Hopkins University Press.

Lucas, Adrienne M. 2010. "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka." *American Economic Journal: Applied Economics* no. 2 (2): 46–71.

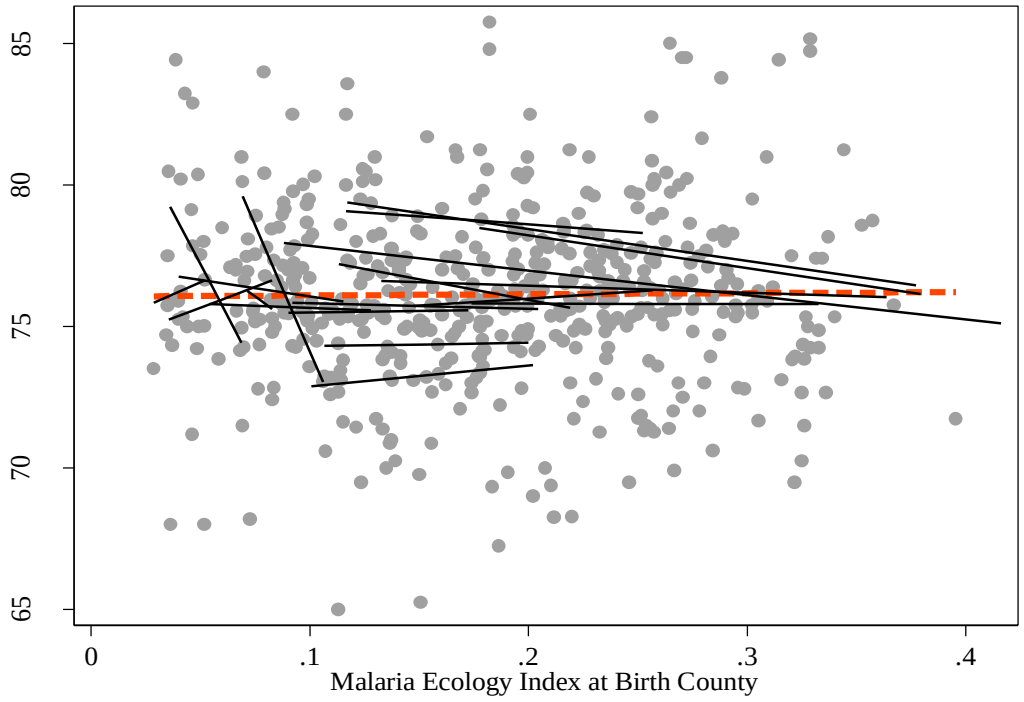
Montez, Jennifer K., and Mark D. Hayward. 2014. "Cumulative Childhood Adversity, Educational Attainment, and Active Life Expectancy among U.S. Adults." *Demography* no. 51: 413-435.

U.S. Census Bureau. 1855. *Mortality Statistics of the Seventh Census of the United States, 1850*. Washington, D.C.: A.O.P. Nicholson Printer.

WHO. 2017. *World Malaria Report 2017*. Geneva: World Health Organization.

Zhu, X., W. Han, W. Xue, Y. Zou, C. Xie, J. Du, and G. Jin. 2016. “The Association between Telomere Length and Cancer Risk in Population Studies.” *Scientific Reports* no. 6: 22243.

Figure 1. Average Age at Death by Malaria Ecology Index at the County of Birth



Note : I clustered 11,006 veterans for baseline estimations to 917 cohorts in terms of year and county of birth. The scatter plot in the figure indicates each cohort's average lifepan and average malaria risk at birth. The dashed line is the linear relationship between the two variables estimated with the pooled data. The solid lines are the linear relationship estimated by state of birth.

Table 1. Estimated Effect of Early-Life Exposure to Malaria on Age at Death

Dependent variable: Age at death

	Baseline estimation					County-of-birth FE	IV estimation		OLS with the sample used for IV estimation
				Baseline 1	Baseline 2		First (Malaria)	Second	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Malaria ecology at the county of birth	0.1877	-0.8426***	-0.9250***	-0.8641***	-0.8373**	-0.7977*		-0.9062	-0.8796**
	(0.1384)	(0.3032)	(0.3390)	(0.3324)	(0.3319)	(0.4170)		(3.4991)	(0.4224)
Ratio of birth county's adult populations from father's birth state or country							0.4173***		
							(0.0307)		
Weak identification Test (F-value)							105.02		
Endogeneity Test (p-value)								0.9939	
Controls:									
Year of birth	Y	Y	Y	Y	Y	Y	Y	Y	Y
State-of-birth FE		Y	Y	Y	Y		Y	Y	Y
Early-life county conditions			Y	Y	Y		Y	Y	Y
Wartime experience				Y	Y	Y	Y	Y	Y
Dummy of chronic conditions by 1899					Y	Y	Y	Y	Y
County-of-birth FE						Y			
Observations	11006	11006	11006	11006	11006	10927	6695	6695	6695

Note: I estimate models (1)-(5) per equation (1), changing the type of control variables. Model (6) uses county-of-birth FEs estimation rather than state-of-birth FEs. Models (7) and (8) are the result of 2SLS estimation, where the ratio of birth county's adult populations from father's birth state or country is used as the instrumental variable for (normalized) malaria ecology index at the county of birth. Model (9) applies the baseline estimation to the sample used for the IV estimation. The size of sample decreased in models (7)-(8) in the process of calculating the instrumental variable. Standard errors are clustered on birth county. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

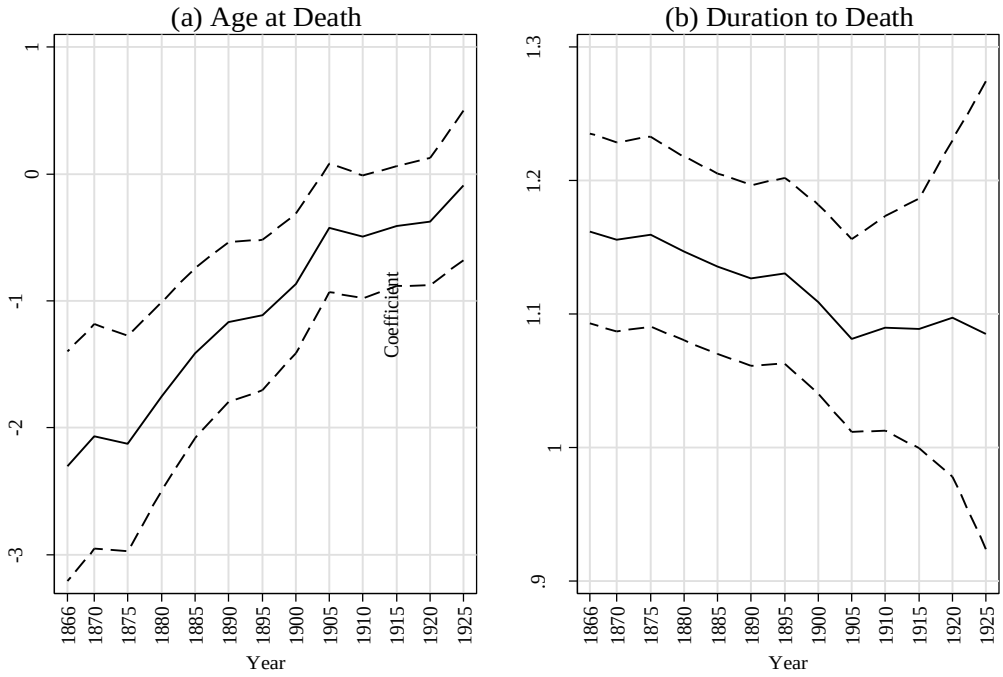
Table 2. Significance of the Exposure to Malaria in Early Life

Dependent variable: Age at death

Sample	1900 links			1860-1900 links				1850-1860-1900 links	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Malaria ecology:									
at the county of birth	-0.8986**	-0.9116**	-1.0241**	-1.2491**	-1.2374**	-1.3453**	-1.4099**	-1.4211**	-1.9568**
	(0.3988)	(0.3987)	(0.4065)	(0.5716)	(0.5724)	(0.6051)	(0.6048)	(0.6845)	(0.9518)
at the county in 1900			0.2017				0.1977		0.2056
			(0.1981)				(0.3135)		(0.3810)
at the county in 1860						0.1883	0.1082		0.5162
						(0.3267)	(0.3588)		(0.5020)
at the county in 1850									0.1228
									(0.8234)
Baseline 2's controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional controls:									
SES in 1900		Y	Y	Y	Y	Y	Y	Y	Y
SES in 1860					Y	Y	Y	Y	Y
SES in 1850									Y
Observations	7592	7592	7592	3894	3894	3894	3894	2621	2621

Note : Each regression utilizes the veterans linked to the 1900, 1860, and/or 1850 census manuscripts, as denoted in the first row. Each contains the control variables and state-of-birth FEs used for model (5) in Table 1. I also control for (normalized) malaria ecology index at the residential county and socioeconomic variables in 1900, 1860, and/or 1850. Standard errors are clustered on birth county. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

Figure 2. Effect of Early-life Exposure to Malaria Risk on Remaining Lifespan by Age



Note : I choose veterans who survived up to a certain year t in 1866-1925, which is denoted in the x axis in the figures above. Then, I run the baseline regressions and plot their estimated coefficients (denoted by solid line) and 90% confidence intervals (denoted by dashed lines) in Figure 2 (a). In Figure 2 (b), I run the Cox proportional hazard regressions with the same control variables as those of the baseline estimation. The dependent variable of the hazard model is the duration to death. I plot their coefficients and 90% confidence intervals in Figure 2 (b).

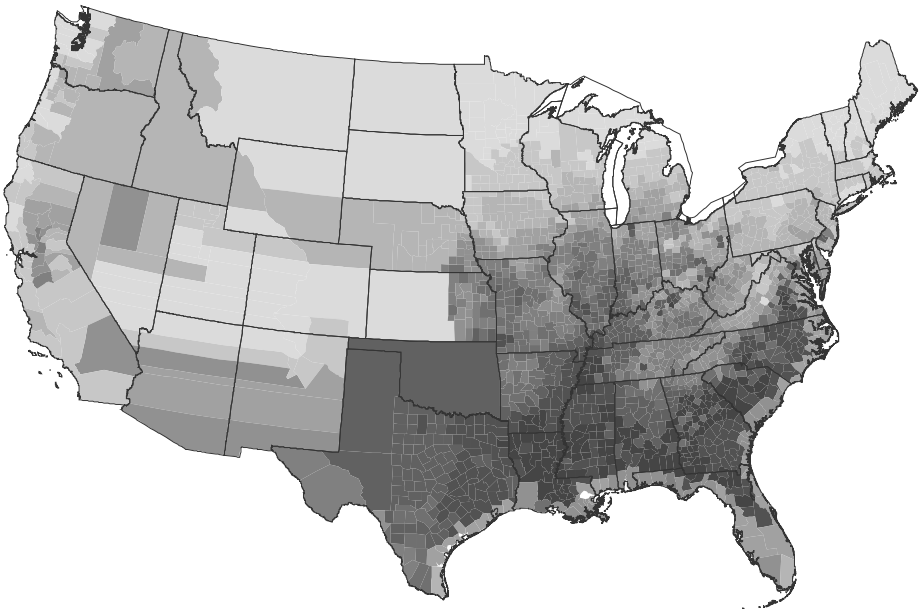
Table 3. Early Exposure to Malaria Risk and Cause of Death

Dependent variables: Dummy of specific cause of death

Sample	All veterans		War survivors		Those who survived by 1900		
	Mean	Baseline 1	Mean	Baseline 2	Mean	Baseline 1	Baseline 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cause of death:							
Diseases of the circulatory system	0.289	-0.0118	0.357	-0.0150	0.421	-0.0200	-0.0212
		(0.0208)		(0.0242)		(0.0284)	(0.0289)
Diseases of the genitourinary system	0.083	-0.0145	0.102	-0.0151	0.125	-0.0193	-0.0155
		(0.0131)		(0.0160)		(0.0197)	(0.0199)
Diseases of the respiratory system	0.103	0.0475***	0.116	0.0576***	0.109	0.0541***	0.0550***
		(0.0139)		(0.0159)		(0.0170)	(0.0169)
Symptoms not elsewhere classified: senility	0.071	-0.0285**	0.074	-0.0439***	0.073	-0.0332**	-0.0342**
		(0.0115)		(0.0130)		(0.0148)	(0.0149)
Neoplasms	0.041	0.0076	0.051	0.0094	0.060	0.0232	0.0247*
		(0.0091)		(0.0113)		(0.0149)	(0.0149)
Certain infectious and parasitic diseases	0.193	-0.0144	0.118	0.0071	0.058	0.0063	0.0044
		(0.0189)		(0.0172)		(0.0158)	(0.0159)
Diseases of the digestive system	0.043	0.0217**	0.052	0.0267**	0.047	0.0272**	0.0260**
		(0.0091)		(0.0109)		(0.0125)	(0.0126)
Diseases of the nervous system	0.040	0.0012	0.047	0.0016	0.043	-0.0070	-0.0085
		(0.0078)		(0.0097)		(0.0120)	(0.0120)
External causes	0.102	0.0184	0.044	0.0015	0.032	0.0016	0.0011
		(0.0135)		(0.0095)		(0.0095)	(0.0094)
Observations		8,631		6,954		4,896	

Note : I run a linear probability regression for each cause of death listed in the left column. So the sample for regressions is limited to those whose cause of death is known. Its key variable is the (normalized) malaria ecology index at the county of birth. Other controls and specification are the same with those of the baseline estimation in column (4) or (5) of Table 1. I report only the estimated coefficient and standard error, clustered on birth county, of the key variable. The regressions in column (2) above use all the available veterans regardless of their year of birth; those in column (4) use the veterans who survived the Civil War; those in columns (6) and (7) use the veterans who survived up to 1900. Columns (1), (3), and (5) report the sample mean of the dependent variable, i.e., the dummy variable of cause of death. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

Figure A1. Estimated Pre-Eradication Malaria Ecology



Note: The index is depicted over the 1860 county boundary. Darker areas are more malarial.

Table A1. Control variables used for the regressions in Section 3

Variables	Description and Sources
Year of birth	Year when the veteran was born
	Early-life conditions: Characteristics of the County of Birth
Crude death rate	Total number of deaths divided by total number of population (ICPSR #2896).
Fertility rate:	Total number of newborns divided by the number of females aged between 15 and 49 (ICPSR #2896).
School enrollment rate	The number of students enrolled in public schools, academies and other schools divided by the number of the population aged between 5 and 14 (ICPSR #2896).
Average farm value	Value of farms divided by farmland in acres (ICPSR #2896)
Slave populations	The number of slaves divided by the size of total populations (ICPSR #2896).
Available transportation	Dummy variables that show whether the county was connected by water transportation or railway (ICPSR #2896).
Dummy for the South	Dummy for being born in the South
Cause-specific mortality rate	From the mortality census manuscripts in 1850, which are available from Ancestry.com and the Center for Population Economics, I counted the number of deaths by nine major infectious diseases for each county: yellow fever, typhoid, smallpox, scarlet fever, pulmonary tuberculosis, pneumonia, measles, diarrhea and cholera. Their mortality rates were estimated by dividing deaths with the number of county populations.
Infant mortality rate	Also estimated using the mortality census manuscripts in 1850.
	SES in 1860 (source: 1860 census records)
Household wealth	Logarithm of the total value of real estate and personal properties possessed by household members.
Urban residence	Dummy for living in urban counties with a population above 10,000.
Literacy	Dummy variable that shows whether the veteran was illiterate or not.
	Wartime Experiences (source: Military records)
Occupation at enlistment	Dummy for showing the type of occupations at enlistment: professional worker, farmer, white collar worker or blue collar worker.
Military rank	Dummy variable that indicates whether their initial rank was private or not.
Wartime diseases and injuries	Dummy variables that show whether the veteran contracted specific infectious diseases or was injured while in military service. The diseases include diarrhea, malaria, typhoid, pneumonia, measles, smallpox, tuberculosis and rheumatic fever.
Prisoner of war	Dummy for the experience of being a POW
Company mortality rate	The number of wartime deaths out of the total number of recruits in the company. This measures the level of wartime stress.
	SES in 1900 (source: 1900 census records)
Marital status	Dummy for being married.
Occupation and urban residence	Dummy variables for showing whether the veteran was a farmer or non-farmer in a rural area.
Home ownership	Dummy variable that shows whether the veteran owned a home or not.