The Socioeconomic Gradient in the Inheritance of Longevity: A Study of American Genealogies

Sok Chul Hong*

Jiwon Park†

March 2016

Abstract

Using American genealogical records on cohorts born in 1845-1884 and their parents, we study the inheritance of longevity across two generations in mid-19th-century America. We find that the level of inheritance was intensified by socioeconomic status in children's early life, which is measured by father's occupational income score. This pattern is estimated more substantial among those born in unhealthy urban states than in rural states. Our study suggests that a better socioeconomic condition helped to avoid the risk of premature death in unhealthy areas so that this strengthened the intergenerational transmission of longevity.

Keywords: Inheritance of Longevity, Socioeconomic Gradient, American Genealogies, Nineteenth-century America
JEL codes: I10, J11, N31

* Associate Professor, Department of Economics, Seoul National University, Email: sokchul.hong@snu.ac.kr

† Student in Master Program, Department of Economics, Sogang University, Email: irene219z@gmail.com

1. Introduction

Researchers have investigated whether long-lived parents have long-lived children since the turn of the 20th century, i.e., the inheritance of longevity. They have found positive correlations between parents' and children's lifespan in various genealogical records complied from different periods, populations and countries. If genetic factors are the unique determinant of an individual's lifespan, the level of the inheritance of longevity would be fixed at the moment of conception. However, a growth body of research has revealed that longevity can be largely modified by individuals' socioeconomic conditions. This strongly suggests that the level of the inheritance of longevity can change depending on socioeconomic conditions.

Some studies have examined the role of socioeconomic conditions in the aspect of the intergenerational transmission of health status. For example, Bhalotra and Rawlings (2013), using the data from the Demographic and Health Survey project, show that children born from shorter (so unhealthy) mothers died in infancy more likely than those from healthy mothers, but the correlation becomes weaker as mother's educational level increases. This provides useful implications that socioeconomic conditions can compensate adverse health status inherited from unhealthy mothers. Similar implications are also found in Currie and Moretti (2007) and Kim et al. (2014).

Longevity is a measure of individuals' long-term outcome, which results from accumulation of health and aging. Therefore, studying the association between socioeconomic conditions and the inheritance of longevity will provide longer-term implications for better understanding the plasticity of aging. But this topic has been rarely explored mainly because genealogical records, which have been frequently used for the study of longevity, generally do not contain individuals' socioeconomic information.

This current study investigates this less explored subject by utilizing an American genealogical sample that consists of 1,459 individuals born in 1845-1884 and their parents. We obtain basic demographic information such as lifespan and gender from their family histories; we measure familial socioeconomic conditions in children's early life with a kind of income index estimated on the basis of fathers' occupations in the 1850-1880 census manuscripts.

The key finding of this study is that the level of the inheritance of longevity strengthened as familial socioeconomic status increased. This is contrary to what recent studies have suggested. We examine and discuss the difference in the historical perspective of nineteenth-century America.

Epidemiological environments had deteriorated throughout the mid-19th century, and so people were largely exposed to a risk of premature death especially due to infectious diseases. As people dwelled under lower-grade environments, the correlation of longevity between generations would be weaker because they died earlier before approaching the lifespan biologically inherited from parents. Therefore, familial socioeconomic conditions would play a key role in preventing premature death and, as its subsequent result, increasing the correlation of longevity. It is also suggested that this role of socioeconomic conditions would be more effective under poorer epidemiological environments. In this study, we provide empirical evidence that supports the hypothesis above.

The main implication of our study is that the direction and scale in which socioeconomic status affects the inheritance of longevity (and probably health) can alter, depending on the prevailing conditions of other longevity determinants like epidemiological environments. Today, many developing countries suffer from infections and inferior sanitation, which cause a high level of premature mortality. Our historical study suggests that the problem can be reduced by enhancing socioeconomic status among individuals and in society.

2. Literature and Background

The intergenerational transmission of longevity has been studied since the turn of the 20th century when genealogy data began to be systematically investigated. Beeton and Pearson (1899), which utilized English genealogies of peerage and landed gentry covering the 17th to 19th centuries, is one of the earliest studies. They estimated strong positive correlations between the ages of death of father and son, and raised the possibility that one's longevity is inherited. Since then, the existence and significance of the inheritance of longevity has been well documented in different periods and countries (Gavrilova and Gavrilov 1999).

Understanding the determinant of an individual's longevity helps us to figure out the mechanism in which parents' and children's lifespan is positively correlated. Genetic factors are the first thing that comes to mind. Scientists have sought specific genetic factors that affect longevity in humans, especially at advanced ages (Hjelmborg et al. 2006). However, human family studies have shown that approximately only 20-30% of the variation in lifespan can be attributed to genetic factors (Herskind et al. 1996). This implies that a majority of variations in longevity are associated with environments including socioeconomic status, health-related behaviors and epidemiological conditions particularly in early life (Christensen and Vaupel 1996).

Economists have emphasized the significance of socioeconomic status. A large number of studies have shown that an individual's income, occupation and education can greatly influence own health status over the course of lifetime and so his/her length of life (Cutler et al. 2006; Cutler et al. 2011). Other studies have reported the transmission of socioeconomic status over generations (Solon 2002). In particular, Currie (2009) shows that health plays a key role in the socioeconomic transmission. Both lines of researches suggest that the inheritability of socioeconomic status can result in the inheritance of health and longevity. Some provides supporting evidence. Case et al. (2002), using contemporary U.S. data, find that a family's long-run average income determines children's health status by delaying the aging process. The health of children from families with lower incomes erodes faster with age, and these children enter adulthood with both lower socioeconomic status and poorer health.

On the other hand, a voluminous body of research has found that early-life exposure to poor epidemiological environments such as hunger, disease, pollution and stress---especially in utero---has adverse impacts on physiology and economic outcomes in adulthood (Almond and Currie 2011; Bleakley 2007; Hong 2013). Some findings in recent studies suggest that such physiological changes may result in an adverse intrauterine environment for the offspring of the mother, leading to physiological changes in the next generation (Currie and Moretti 2007; Lee 2014). Scientists discuss its mechanism in the aspect of epigenetic modification, which says that a stimulus or insult received during critical periods during critical periods of growth and development may cause heritable change in gene activity and expression (Feil and Fraga 2012).

The intergenerational studies above strongly propose that improvement in socioeconomic status and epidemiological environments can change the impact of the transmission of health and longevity across generations. A few studies have explored the proposition. Bhalotra and Rawlings (2013) is a recent study that emphasizes the effects of improvements in the early-life socioeconomic conditions. They utilize individual survey data on 600,000 mothers and their children in 38 developing countries to estimate the intergenerational transmission of health. They estimate that maternal stature (measure of maternal health) has a substantial negative correlation with infant mortality (inverse measure of children's health) across countries, i.e., the inheritance of health status. The main finding of the study is that a higher level of mother's educational attainment can limit the extent to which children born to shorter women are condemned to death in infancy. Currie and Moretti (2007) and Kim et al. (2014) similarly find that the intergenerational correlations of health are weaker in areas with higher average income.

Although evidence on the association between the inheritance of health and socioeconomic status is not enough, studies that investigate the role of socioeconomic status on the inheritance of longevity are very rare. This is mainly because longevity studies require historical data covering much longer periods and such data like genealogical records generally do not contain individuals' socioeconomic variables. Longevity is a long-term outcome that is determined by lifetime interactions among health, socioeconomic status and environments. So understanding the socioeconomic gradient in the inheritance of longevity will provide insights into the plasticity of aging. This current study helps to fill in the unexplored subject.

This study also has advantage in that it explores the socioeconomic gradient in the inheritance from the historical framework of mid-19th-century America. The period is characterized by infectious diseases, poor sanitation, lack of social and public health programs and so on (Floud et al. 2011). People were at a high risk of premature death (Preston and Haines 1991; Fogel 2004). Costa (2015) argues that the changing relationship among health, income and the environments dynamically led to the long-term health transition in the United States throughout the centuries. Thus, the role of socioeconomic status and its impact on the inheritance of health and longevity can be different with what modern studies have found.

3. Genealogical Data and Socioeconomic Variable

This study uses an intergenerational database: the Intergenerationally-Linked Aging Sample (ILAS). The whole sample was collected from printed American family histories and genealogies compiled for 39 families (lineages) composed of about 15,000 households. The ILAS data, which is one of the richest collections of U.S. family histories in terms of sample size and years covered, includes 118,062 individuals with birth dates ranging from 1577 to 1983. The core data file based on family history records provide each individual's identification number, name, gender, years of birth, marriage and death, places (state, county, and town) of birth and death, and identification numbers of parents, grandparents and family. In addition, 27% of observations are linked to at least one of the 1850-1880 and 1900-1910 US federal manuscript censuses.[1] The census file provides rich information on individual and familial socioeconomic status and environmental conditions. Therefore, the ILAS data set can compensate for the weakness of other databases used for human-longevity studies.[2]

The ILAS data were collected and first studied by Clayne L. Pope. Pope's (1992) pioneering work, which used a smaller version of the ILAS data, examined long-term U.S. mortality trends.[3] In the pilot study, he also argues that the ILAS data do not have any critical biases and shows the usefulness of family histories for studying human longevity. However, as Pope (1992) pointed out, we should carefully consider potential bias due to missing variables and attrition. (We discuss the representativeness of the ILAS in Appendix A.)

First, the missing-variable problem is caused by the fact that death years in the ILAS data are frequently missing particularly for childhood deaths. In general, those who died in infancy were generally included in the family history (most often with a birth year), but often with no death year. Thus, death years are about one third as common in the ILAS data as are birth years. This is mainly because of high infant mortality in the past and partly because the dispersal of families and married women's name changes make death years harder to find. To fix bias due to missing information of deaths at young ages, Pope (1992) suggested limiting the sample to cohorts who were alive at adult ages (at least age 20). The trend of life expectancy at age 20 estimated under the constraint of Pope (1992) is similar with those estimated by other studies like Kunze (1979),

---

[1] The 1890 census manuscripts were destroyed by fire in 1921.

[2] Another weakness of existing genealogical records is that they represent specific populations such as the nobility (e.g., British Peerage Genealogy DB, 1603-1959; Qing Nobility DB, 1640-1911) or specific localities (e.g., Family-Reconstruction Data on French Village of Arthez d'Asson, 1744-1975); do not cover enough generations to study the transmission of longevity (e.g., Hauge-Harvald DB on Elderly Danish Twins, 1870-1930); or, focus on long-lived persons (e.g., Swedish Centenarian DB, about 5,000 Belgian centenarians born in 1870-1897) (Gavrilova and Gavrilov 1999). Compared with them, the ILAS more presents general populations and covers longer periods and broader areas in the United States.

[3] In the pilot study, he drew 49,419 individuals from 23 different family histories.

and particularly well reflect the marked decline of life expectancy in the antebellum period noted by Fogel (1986).

Second, the attrition problem is related with sample distribution of birth years: 40% of the ILAS observations were born after 1890. Although the number of people in each completed generation increases, the rate of increase is below the expected rate for most families because some individuals in each generation are not followed in the basic genealogical records, which eliminates their descendants from the book. Moreover, those born near the completion of the book are at risk of dying up until the complier stopped collecting data. Thus, the ratio of individuals whose death years are known is low for the cohorts born after 1890. To fix the attrition bias, Pope (1992) suggested limiting the sample to cohorts who were born before 1890.

The current study examines the inheritance of longevity across generations, and so is exposed to two problems above. Following the methods suggested by Pope (1992), we limit the sample to those i) who survived at least up to age 20, ii) who were born before 1885, and iii) whose parents' birth and death years are known.[4]

However, we impose additional constraints on the selection of sample according to the purpose of analysis. We particularly estimate fathers' occupation-based income index---so-called occupational income score---to measure socioeconomic conditions in children's early life, utilizing the information on paternal occupations found in the census manuscripts. Thus, we select children whose fathers are linked to the census records of the decade when children were born. The census manuscripts are available from the 1850 federal census; we define decade *t* as the years from *t*-5 to *t*+4 in this study. Therefore, the final sample includes 1,458 children born in 1845-1884 and their parents.

On the other hand, we estimate the occupational income score by merging fathers' occupations with a list of income score by occupation. We constructed the list from the 1850-1880 IPUMS data, which estimate the income score by occupation on the basis of 1950 information. Thus, the occupational income score is a kind variable of real income. We report details on data source and summary statistics in Appendix B.

[4] We also limit parents sample into those who survived at least up to age 20. But most parents satisfy the condition.

4. The Intergenerational Transmission of Longevity and Socioeconomic Gradient

With the selected sample, we first estimate the level of the inheritance of longevity and its statistical significance using the following equation.

$$L_{ipst} = \alpha + \beta_l L_p + \beta_o O_p + X_i \Gamma + \delta_s + \delta_t + u_{sib} + \varepsilon_{ipst} \quad (1)$$

In equation (1), L_{ipst} denotes the age at death of child i who was born from parent p in state s and decade t. The key explanatory variable of children's longevity is parental lifespan denoted by L_p. O_p is the variable of fathers' occupational income score in the decade of child's birth. In addition to constant term, we control for some available variables that can be related to longevity (X_i) including parental age at the child's birth, child's gender, number of siblings, and the dummy variable of being born in large cities with 25,000 populations. In particular, the dummy variable partially captures the effect of early-life epidemiological conditions.

The regressions also contains children's state-of-birth fixed effects (δ_s) to control for time-invariant state-specific factors; the use of children's decade-of-birth fixed effects (δ_t) will capture time trend and decade-specific determinants of longevity. Finally, we add sibling random effects (u_{sib}) to control for family-specific factors such as genotype and frailty that siblings share. In the specification, we assume that sibling-specific investments are uncorrelated with the genetic endowment even though, on average, parental investment may reinforce or compensate early-life differences (cf. Almond and Mazumder 2013; Conti et al. 2011).

Table 1 reports only the estimated coefficients of parental lifespan. The level of the inheritance of longevity is estimated similarly in models (1) and (2), even controlling for two fixed effects and sibling random effects.[5] In fact, controls other than parental lifespan do not significantly account for the variation in children's longevity. Model (2) says that a one-year increase in mother's lifespan led to about 0.19-year increase in children's longevity in panel A, and the corresponding marginal effect of father's lifespan is estimated by about 0.13 years in panel B. Panel C, which includes maternal and paternal factors together, produces similar coefficients

[5] The Breusch-Pagan Lagrange multiplier test in model (2) suggests that the use of sibling random effect is valid. We use the model (2) as the baseline model in the remaining analyses.

and still shows that the inheritance of longevity from mother is stronger than that from father. It is known that mothers can have stronger influence on the offspring because of specific inheritance of mitochondrial DNA (Falconer 1989). But the strong link would be also related with strong maternal-child interactions in utero development and during growing years. But the mechanism is still unknown.

[Table 1 Here]

In model (3), we estimate the intergenerational longevity elasticity by taking the logarithms of parental and children's lifespan. The result is similar with the marginal effects in model (2): about 0.19 for mother-child elasticity and about 0.14 for father-child elasticity. It is intriguing to know that the estimates above are similar with the intergenerational elasticity of socioeconomic status estimated by other studies. Although reliable estimates for 19th-century America are not available, some studies for early-20th-century America show that the intergenerational elasticity of income and education is around 0.2 (Lee and Solon 2009; Chetty et al. 2014; Hertz et al. 2007).[6] This suggests that longevity can be considered as a long-term outcome, which is largely influenced by socioeconomic status in lifetime.

Finally, we estimate the impact of the inheritance of longevity for sons and daughters, separately, in models (5) and (6). In both, mother's lifespan has more influence on children's longevity. But the longevity of sons is more closely associated with father's lifespan, and that of daughters is more influenced by mother's lifespan. This pattern is also found in some other studies including Parman (2012) that is based on death-certificate data consisting of those from North Carolina who died in 1934-1974 and their parents who died after 1909. This seems to result from various factors including parents-children interaction pattern from in-utero to later life and genetic factors. But we do not find any clear scientific evidence.

Now, we investigate how the gradient of the inheritance of longevity was different by the level of familial socioeconomic conditions around children's birth. The emphasis of the conditions

[6] The studies show that the elasticities increased over time and they are around 0.3-0.5 in modern years.

around children's birth reflects the findings of existing studies on the significance of in-utero or early-life conditions in determining lifetime health and human capital as discussed in Section 2.[7]

For the analysis, we interact socioeconomic status (O_p), which is measured by the logarithm of fathers' occupational income score, with parental lifespan (L_p) as follows:

$$L_{ipst} = \alpha + \beta_l L_p + \beta_{l\times o}\left[L_p - \overline{L_p}\right]\left[O_p - \overline{O_p}\right] + \beta_o O_p + X_i\Gamma + \delta_s + \delta_t + u_{sib} + \varepsilon_{ipst} \qquad (2)$$

We use demeaned variables when we construct the interaction terms.[8] Other specifications are the same as model (2) of Table 1.

The estimated coefficients of key variables are reported in Table 2's model (1). The main finding here is that the inheritance of longevity becomes significantly stronger as fathers' occupational income score increases, although the income score itself does not have significant impact on children's longevity. This interaction effect is found in both maternal and paternal sides.

[Table 2 Here]

The socioeconomic gradient in the level of the inheritance of longevity is substantial. The coefficient in panel A suggests that as father's income score increases by 10% from its mean, the influence of maternal lifespan on children increases by 0.0223 in terms of coefficient or by 11.7% from the average marginal effect (i.e, β_l). The magnitude from father's lifespan in panel B is a bit greater as a 16.5% increase in the slope for a 10% increase in income score. This may suggest that the longevity link between mother and child is more biological than the father-child longevity link.

One may argue that father's occupational income score may not exactly capture familial socioeconomic conditions in children's early life. Thus, we alternatively estimate father's occupational income score as lifetime average and use it in models (3) and (4). For the variable, we utilize father's census linkages to the 1850-1910 census manuscripts, merge all the occupations

7 In addition, information such as place of birth and socioeconomic/epidemiological conditions is available for many observations in the ILAS data set. But it is hard to find out place of residence and environmental conditions for their later lives.

8 In equation (2), $\overline{L_p}$ and $\overline{O_p}$ denote the sample mean of L_p and O_p, respectively.

at ages 20-59 found in multiple censuses with the IPUMS income-score information, and calculate the average of income scores across the censuses. The use of alternative socioeconomic variable little changes the result and implication.

On the other hand, we report the estimation result for income score and the dummy of being born in large cities in Table 2. Although all the coefficients are not statistically significant, their signs suggest that socioeconomic status helped to increase children's longevity and being born in large cities caused early deaths.

5. The Socioeconomic Gradient by Epidemiological Environments

The main finding above is that familial socioeconomic status in early life reinforces the level of the inheritance of longevity. As discussed in Section 2, modern studies show that the inheritance can be weaker when socioeconomic status compensates the adverse health inheritance from unhealthy parents. In this section, we examine the difference in the historical perspective of 19th-century America.

The most influencing determinant of longevity in the nineteenth century is epidemiological environments. Throughout the mid-nineteenth century, the conditions of infectious diseases and sanitation had seriously deteriorated. Consequently, the major indicators of public health such as infant mortality rate, life expectancy, and average adult height continued to worsen during the periods (Floud et al. 2011). It is obvious that people---especially infants---were largely exposed to the risk of premature death over the course of lifetime, and many people unexpectedly died earlier prior to ages genetically programmed. This would be more frequent in unhealthy areas filled with infectious diseases. Longevity link or genetic cohesion across generations would be weak in the areas. This hypothesis is tested and approved by Lauderdale and Rathouz's (1999) study on Union Army veterans' sample. They show that the correlation of height between brothers (a measure of genetic cohesion) decreases with increasing county population (a measure of unhealthy environment).

On the other hand, historical studies show that nineteenth-century Americans from families with higher socioeconomic status significantly had higher nutritional status, better health outcomes and longer life than otherwise (Preston and Haines 1991). Considering the significance of epidemiological environments, the grave role of familial socioeconomic status for health and longevity would be to prevent or avoid the risk of premature death due to infections, as suggested by Warren et al. (2012).

Putting the characteristics of nineteenth-century America together, we propose two hypotheses. First, the inheritance of longevity would be weak in unhealthy areas since people in the areas died earlier than expected. Even though they did not die in infancy, early-life exposure to infections would speed up the process of aging in adult and shorten the length of life. Second, familial socioeconomic status would be effective in avoiding the risk of premature death in unhealthy areas, but this would work less effectively in healthy areas. Accordingly, as socioeconomic status improves, the positive correlation of longevity between generations would more significantly increase in unhealthy areas than in areas with good environments.

We test the above hypotheses by comparing the socioeconomic gradient in the level of the inheritance of longevity between healthy and unhealthy areas. For the division of two areas, we use three state-level variables measured for the state and decade of children's birth: population density, ratio of state populations in cities with over 25,000, and crude death rate. It is rational to assume that children born in dense, urban or high-mortality states more likely suffered from unhealthy environments (Cain and Hong 2009). Equation (2) is estimated for two subsamples born in healthy and unhealthy states, respectively.

The results in Table 3 well approve the hypotheses proposed above. First, for three measures of state epidemiological environments, the link of longevity between generations is estimated weaker in unhealthy areas than in healthy areas. Second, the coefficient of the socioeconomic gradient is estimated significantly positive in unhealthy states, but it is statistically insignificant and small in healthy states. This implies that familial socioeconomic status reinforced the inheritance of longevity by lowering the probability of premature death.

[Table 3 Here]

The role of socioeconomic status looks great. For example, the coefficients in models (1) and (2), a 10%-increase of father's occupational income score increases the slope of the mother-child longevity inheritance by 5% in low dense states, but by 17% in high dense states. The impact on the slope of the father-child inheritance is more substantial: 5% versus 26%.

Finally, we take one more step to confirm the association between the role of socioeconomic status and the level of epidemiological environments. The risk of premature death is the key factor that causes the association of two components. Thus, the association can be different across the groups characterized by the risk of premature death. For example, the risk was different across decades in the nineteenth century. Many studies say that epidemiological environments became worse from the 1850s to the 1880s mainly due to urbanization. This is clearly evidence in the trends of infant mortality rate and life expectancy.

In Table 4's panel A, we do the same analysis as done in models (1) and (2) of Table 3 for two cohorts born in 1845-1864 and 1865-1884, respectively. The later cohort more suffered from unhealthy environments in early life. Two things are worthy to discuss in the table. First, average effect of maternal lifespan on children's longevity is estimated much lower among 1845-1864 cohorts in model (1) (i.e., the low-risk group) than among the high-risk group in model (4). This is consistent with Lauderdale and Rathouz (1999) and what we discussed in Table 3. Second, the difference in the socioeconomic gradient by state population density is clearly found among the high-risk cohorts, but it is not among the low-risk cohorts. This strongly suggests that socioeconomic status played an effective role in preventing the risk of premature death and so reinforcing the link of longevity only when epidemiological environments were unhealthy.

[Table 4 Here]

Panel B compares sons and daughters. In nineteenth-century America, males were exposed to a higher level of premature-death risk over the course of lifetime. This is supported by the vital statistics that men's infant mortality rate and life expectancy were inferior to those of women. The result is very similar with that of panel A. The average link of longevity is found lower in the high-risk group. Only the group also shows a statistically significant difference in the socioeconomic gradient by state population density.

In panel C, two groups are constructed more artificially. We contain children who died before age 20 in the sample, who are thought to have died prematurely. Thus, this group is considered to have exposed to a high risk of premature death. Whereas we select only children and parents who survived up to age 50 for the low-risk group. Although the implication from the difference in average marginal effect is opposite, the disparity in the socioeconomic gradient by epidemiological environments is clearly found only among the high-risk group.

6. Concluding Remarks

In this study, we have explored the intergenerational transmission of longevity using an American genealogical sample that consists of individuals born in 1845-1884 and their parents. We particularly examined how familial socioeconomic status---measured by father's occupational income score around children's birth---affected the level of the transmission of longevity, which has been rarely studied in previous researches.

The key findings are summarized as follow: First, a significant positive correlation in lifespan between parents and children is estimated. Its magnitude is very comparable with other estimates of intergenerational mobility based on income and educational attainment. Second, the socioeconomic gradient in the inheritance of longevity is significantly estimated. In other words, the positive correlation of longevity was intensified as family had a higher-level of socioeconomic status. Third, we hypothesized that socioeconomic status would play a key role in preventing or avoiding the risk of premature death and that the socioeconomic gradient would be more substantial in the region or group in which the risk was more prevailing. We test and approve the hypothesis by estimating the socioeconomic gradient across various subsamples.

The socioeconomic gradient in the inheritance of health status has been reported in some studies using modern data. Those studies generally show that the level of the inheritance of health is more substantially found among the group with lower socioeconomic conditions. The direction of the socioeconomic gradient in our study is apposite to that in the studies above. However, we should emphasize that the socioeconomic role of improving the intergenerational link is consistent in both studies. People in mid-nineteenth-century America were largely exposed to the high risk of premature death due to infectious diseases and poor sanitation. Therefore, the longevity correlation would be stronger among families with better socioeconomic conditions because they could effectively avoid the risk of premature death. Today, socioeconomic status is thought to compensate the low health capital inherited from unhealthy parents.

The main implication of this study is closely related with developing countries. Infectious diseases, poor sanitation and various environmental problems like pollution prevails among many developing countries today. When the countries do not have adequate social programs to resolve such problems, individuals' and familial socioeconomic status would be the most effective measure to avoid the risk of mortality especially among children. This study provides historical evidence showing how the improvement in socioeconomic status can enhance lifespan and health in the intergenerational perspective in developing countries.

Appendix A. Representativeness of the ILAS

This section investigates whether the ILAS sample is representative of the US population in terms of longevity and socioeconomic status. Table A1 presents the summary statistics for key variables obtained from census linkage and adds the corresponding statistics calculated from 1870 IPUMS. The IPUMS data consist of a 1% sample drawn from the 1870 census manuscripts. The comparison therefore allows us to investigate the representativeness of the ILAS. In Panel A, we compare the characteristics of males aged 20-50 in 1870 (born 1820-1850). Average value of real property is a bit higher in ILAS than IPUMS, but those in IPUMS have a bit more property wealth. Total wealth is not statistically, significantly different. Difference in the fraction of farmers and illiteracy rates are small. In addition, both samples have a similar regional 1870 residential distribution. Panel B compares the characteristics of the children's cohort for those aged 0-20 in 1870 (born 1850-1870). Although IPUMS shows a higher rate of school attendance than ILAS, a t-test rejects the statistical significance of this difference. The comparison in Table 3 suggests that the ILAS data are sufficiently representative with the US population c.1870 in terms of SES and regional distribution.

[Table A1 Here]

On the other hand, we estimated the (period) life expectancy at age 20 (i.e., e20) for each year from the ILAS data set. Figure A1 shows its 10-year moving average in 1820-1890, and also includes e20 estimated in other studies. The estimate from the ILAS data has a declining trend until the 1860s for both men and women, which is the marked decline in the antebellum period. It increases by the 1870s and then becomes stable. Although some disparities in the level of e20 across the studies are observed, the estimate from the ILAS data is consistent with other existing historical estimates.

[Figure A1 Here]

Appendix B. Summary Statistics

The data we used in this study come from three different sources; ILAS, IPUMS and ICPSR No. 2986 that digitized historical census publications. We used 1850-80 IPUMS data to construct father's occupational income score at the child's birth decade in ILAS. Occupational income score is based on the occupation information recoded into the 1950 classification. This indicates the median total income—in hundreds of dollars—of the personas in each occupation in 1950. So occupational income score provides a continuous measure of occupations, according to the economic rewards enjoyed by people working at them in 1950. We take three variables—population density, percent of urban population, crude death rate---from ICPSR No. 2986 to measure the epidemiological environment of the state where children were born.

The summary statistics of individual-level data is in Table B.1. We divide the observations into two groups according to age at death. The mean longevity of children in the sample is 64.5. The mean longevity of parents is bigger, 67.1 and 71.6 for mother and father, respectively. This is because the parents survived to have children so they are less at the early death risk than their children. Each child had 7.4 siblings (including him/herself) in average. There is substantial dispersion in father's occupational income score at child's birth, from 6 to 80 while the average is 17.81

[Table B1 Here]

When comparing the two separate groups, there is no significant difference between two groups except for the longevity. The average longevity of children who died before age of 50 is 35.7 while 72.8 for those who survived up to 50. The differences of parental longevity between two groups are significant. Considering that the mean longevity of parents of children who survived up to 50 is bigger, we can expect the positive correlation between parents' and children's longevity.

References

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives, 25*(3), 153-172.

Almond, D., & Mazumder, B. (2013). Fetal origins and parental responses. *Annual Review of Economics, 5*, 37-56.

Beeton, M., & Pearson, K. (1899). Data for the problem of evolution in man. II. A first study of the inheritance of longevity and the selective death-rate in man. *Proceedings of the Royal Society of London, 65*(413-422), 290-305.

Bhalotra, S., & Rawlings, S. (2013). Gradients of the intergenerational transmission of health in developing countries. *Review of Economics and Statistics, 95*(02), 660-672.

Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American South. *Quarterly Journal of Economics, 122*(1), 73-117.

Cain, L., & Hong, S. C. (2009). Survival in 19th century cities: The larger the city, the smaller your chances. *Explorations in Economic History, 46*(4), 450-463.

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *The American Economic Review, 92*(5), 1308-1334.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics,*

Christensen, K., & Vaupel, J. W. (1996). Determinants of longevity: Genetic, environmental and medical factors. *Journal of Internal Medicine, 240*(6), 333-341.

Conti, G., Heckman, J. J., Yi, J., & Zhang, J. (2010). Early health shocks, parental responses, and child outcomes. Unpublished Manuscript, Department of Economics, University of Hong Kong, Costa, D. (2015). Health and the economy in the united states, from 1750 to the present. *Journal of Economic Literature, 53*(3), 503-570.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature, 47*(1), 87-122.

Currie, J., & Moretti, E. (2007). Biology as destiny? Short-and Long-Run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics, 25*(2), 231-264.

Cutler, D., Deaton, A., & Lleras-Muney, A. (2006). The determinants of mortality. *The Journal of Economic Perspectives, 20*(3), 97-120.

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2011). *Socioeconomic status and health: Dimensions and mechanisms.* Oxford University Press.

Falconer, D. S. (1989). *Introduction to quantitative genetics.* Longman, Scientific & Technical.

Feil, R., & Fraga, M. F. (2012). Epigenetics and the environment: Emerging patterns and implications. *Nature Reviews Genetics, 13*(2), 97-109.

Floud, R., Fogel, R. W., Harris, B., & Hong, S. C. (2011). *The changing body: Health, nutrition, and human development in the western world since 1700* Cambridge University Press.

Fogel, R. W. (1986). Nutrition and the decline in mortality since 1700: Some preliminary findings. *Long-term factors in American economic growth* (pp. 439-556) University of Chicago Press.

Fogel, R. W. (2004). *The escape from hunger and premature death, 1700-2100: Europe, America, and the third world.* Cambridge University Press.

Gavrilova, N. S., & Gavrilov, L. A. (1999). Data resources for biodemographic studies on familial clustering of human longevity. *Demographic Research, 1*(4), 1-48.

Glied, S., Smith, P. C., Cutler, D. M., Adriana Lleras-Muney, & Vogl, T. *Socioeconomic status and health: Dimensions and mechanisms.* Oxford University Press.

Herskind, A. M., McGue, M., Holm, N. V., Sörensen, T. I., Harvald, B., & Vaupel, J. W. (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics, 97*(3), 319-323.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2007). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy, 7*(2)

Hjelmborg, J., Iachine, I., Skytthe, A., Vaupel, J. W., McGue, M., Koskenvuo, M., et al. (2006). Genetic influence on human lifespan and longevity. *Human Genetics, 119*(3), 312-321.

Hong, S.C. (2013). Malaria: An Early Indicator of Later Disease and Work Level. *Journal of Health* Economics, 32(3), 612-632.

Kim, Y., Sikoki, B., Strauss, J., & Witoelar, F. (2014). Intergenerational correlations of health among older adults: Empirical evidence from Indonesia. *The Journal of the Economics of Ageing,*

Kunze, K. (1979). The effects of age composition and changes in vital rates on nineteenth century population estimates from new data. Ph.D. dissertation, University of Utah.

Lauderdale, D. S., & Rathouz, P. J. (1999). Evidence of environmental suppression of familial resemblance: Height among US civil war brothers. *Annals of Human Biology, 26*(5), 413-426.

Lee, C. (2014). In utero exposure to the Korean War and its long-term effects on socioeconomic and health outcomes. *Journal of Health Economics, 33*, 76-93.

Lee, C., & Solon, G. (2009). Trends in intergenerational income mobility. *The Review of Economics and Statistics, 91*(4), 766-772.

Parman, J. (2012). Gender and Intergenerational Mobility: Using Health Outcomes to Compare Intergenerational Mobility across Gender and Over Time. Working Paper, The College of William and Marry.

Pope, C. L. (1992). Adult mortality in America before 1900: A view from family histories. *Strategic factors in nineteenth century American economic history: A volume to honor Robert W. Fogel* (pp. 267-296) University of Chicago Press.

Preston, S. H., & Haines, M. R. (1991). *Fatal years: Child mortality in late nineteenth-century America.* National Bureau of Economic Research, Inc.

Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *The Journal of Economic Perspectives, 16*(3), 59-66.

Warren, J. R., Knies, L., Haas, S., & Hernandez, E. M. (2012). The impact of childhood sickness on adult socioeconomic outcomes: Evidence from late 19th century America. *Social Science & Medicine, 75*(8), 1531-1538.

Table 1. The Level of the Inheritance of Longevity

Dependent variable: Children's age at death

	(1)	(2)	(3)	(4)	(5)
Estimation model	No other controls	Full controls (baseline)	Measuring elasticity	Using model (2)	
				Sons	Daughters
	Panel A: Only maternal factors				
Mother's age at death	0.1852***	0.1957***	0.1910***	0.1510***	0.2330***
	(0.0294)	(0.0324)	(0.0373)	(0.0403)	(0.0526)
Sample size	1459	1459	1459	868	591
R-squared	0.028	0.087	0.078	0.090	0.139
P-value of LM test for REs		0.035	0.073	0.047	0.161
	Panel B: Only paternal factors				
Father's age at death	0.1169**	0.1296**	0.1338**	0.1219*	0.1016
	(0.0374)	(0.0429)	(0.0494)	(0.0515)	(0.0645)
Sample size	1459	1459	1459	868	591
R-squared	0.007	0.066	0.063	0.077	0.114
P-value of LM test for REs		0.002	0.016	0.012	0.271
	Panel C: Maternal and paternal factors together				
Mother's age at death	0.1790***	0.1926***	0.1882***	0.1494***	0.2270***
	(0.0296)	(0.0324)	(0.0375)	(0.0404)	(0.0527)
Father's age at death	0.1002**	0.1207**	0.1257*	0.1192*	0.0975
	(0.0373)	(0.0421)	(0.0495)	(0.0507)	(0.0635)
Sample size	1459	1459	1459	868	591
R-squared	0.033	0.093	0.084	0.090	0.145
P-value of LM test for REs		0.036	0.084	0.050	0.142

Note: We estimate the effect of parental age at death on children's longevity, per equation (1). All models use 1,459 observations who were born in 1845-84 and whose fathers are linked to the census records in the decade of children's birth. Panel A controls for maternal factors only such as mother's age at death and mother's age at the birth of her child; Panel B controls for paternal factors only; Panel C includes maternal and paternal factors together. Model (1) includes the parental age at death and constant term only. We control for all the covariates, two fixed effects and sibling random effect, which are discussed in text, in model (2). The p-values of the Breusch-Pagan Lagrange multiplier tests suggest that the use of sibling random effects is valid. We estimate the intergenerational elasticity of lifespan in model (3) by taking logarithms of parental and children's lifespan. Models (4) and (5) use the same specification with model (2), but we separate sons and daughters in each regression. We report only the coefficients of parental age at death in the table; their robust standard errors are reported in parentheses. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

Table 2. The SES Gradient in the Level of the Inheritance of Longevity

Dependent variable: Children's age at death

	(1)	(2)	(3)	(4)
Father's occupational income score	Value in the decade of children's birth		Average value across the census records	
Parent	Mother	Father	Mother	Father
Parental lifespan [L_p]	0.1933***	0.1306***	0.1936***	0.1323***
	(0.0320)	(0.0421)	(0.0320)	(0.0423)
$L_p \times I_p$	0.0226***	0.0215**	0.0229***	0.0229*
	(0.0069)	(0.0107)	(0.0083)	(0.0117)
Father's income score [I_p]	0.1306	0.1735	0.0710	0.0884
	(0.1460)	(0.1567)	(0.1733)	(0.1864)
Dummy of large cities	-3.5257	-3.2181	-3.8394	-2.9429
	(3.1506)	(3.1414)	(3.1363)	(3.1223)
Sample size	1459	1459	1462	1462

Note: We estimate how the gradient of the inheritance of longevity was different by the level of familial socioeconomic conditions around children's birth, per equation (2). We use demeaned variables when we construct interaction terms. Other specifications are the same as those used in model (2) of Table 1. In models (1) and (2), we use father's occupational income score around children's birth as a measure of familial socioeconomic status; models (3) and (4) use average value of father's income score found in the 1850-1910 census records. Note that the variable of father's income score is calculated in logarithm and multiplied by 10 for better interpretation. We consider the effect of mother's lifespan in models (1) and (3), and father's lifespan in (2) and (4). Robust standard errors are reported in parentheses. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

Table 3. The Socioeconomic Gradient by the Level of Epidemiological Environments

Dependent variable: Children's age at death

	(1)	(2)	(3)	(4)	(5)	(6)
Epidemiological environments:						
Measure	Population density		% of urban populations		Crude death rate	
Level	Low	High	Low	High	Low	High
	Panel A: Effect of mother's life span					
Parental lifespan [L_p]	0.2086***	0.1748***	0.2133***	0.1299**	0.2487***	0.1448***
	(0.0443)	(0.0450)	(0.0372)	(0.0582)	(0.0476)	(0.0421)
$L_p \times I_p$	0.0105	0.0300**	0.0100	0.0310**	0.0131	0.0241**
	(0.0077)	(0.0130)	(0.0079)	(0.0145)	(0.0112)	(0.0096)
Father's income score [I_p]	-0.2092	0.3691	-0.0785	0.1007	-0.1107	0.3363
	(0.1659)	(0.2529)	(0.1809)	(0.2781)	(0.2022)	(0.2089)
Dummy of large cities	-2.8566	-4.5158	-11.3451***	-2.4818	-12.6077**	-1.4558
	(6.2042)	(3.2917)	(2.7447)	(2.9631)	(5.8984)	(3.1512)
Sample size	666	792	965	493	619	837
	Panel B: Effect of father's life span					
Parental lifespan [L_p]	0.1469**	0.1162**	0.1242***	0.0928	0.1939***	0.0911*
	(0.0628)	(0.0565)	(0.0482)	(0.0735)	(0.0663)	(0.0486)
$L_p \times I_p$	0.0074	0.0303**	0.0104	0.0420**	0.0108	0.0309*
	(0.0134)	(0.0151)	(0.0119)	(0.0171)	(0.0148)	(0.0171)
Father's income score [I_p]	-0.2507	0.5727**	-0.1449	0.3987*	-0.0576	0.3869*
	(0.2024)	(0.2339)	(0.1898)	(0.2177)	(0.2314)	(0.2103)
Dummy of large cities	-3.0954	-4.6666	-11.8823***	-3.2270	-12.2893**	-1.1543
	(6.5224)	(3.2495)	(3.3366)	(2.9194)	(6.2442)	(3.2494)
Sample size	666	792	965	493	619	837

Note : We estimate equation (2) for different groups clustered according to state-level environmental measures. For the division, we use three state-level variables measured for the state and decade of child birth: population density, percent of urban populations, crude death rate per 100,000 population. The cut-off value is the average of each variable. The low level denotes healthy conditions and vice versa. Panel A and B consider the effect of mother's and father's lifespan, respectively. Other specifications are the same with those of models (1) and (2) of Table 2, which use father's occupational income score around children's birth as a measure of familial socioeconomic status. Robust standard errors are reported in parentheses. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

Table 4. The Socioeconomic Gradient by the Risk of Premature Death

Dependent variable: Children's age at death

	(1)	(2)	(3)	(4)	(5)	(6)
Premature-death risk	High risk			Low risk		
State population density	All	Low	High	All	Low	High
Panel A: Birth Decade	Birth decade = 1870 & 1880			Birth decade = 1850 & 1860		
Maternal lifespan	0.1224***	0.1920***	0.0452	0.2407***	0.2420***	0.2384***
	(0.0437)	(0.0554)	(0.0654)	(0.0467)	(0.0689)	(0.0622)
Maternal lifespan × Income score	0.0344***	0.0090	0.0496***	0.0080	0.0099	0.0021
	(0.0107)	(0.0132)	(0.0188)	(0.0100)	(0.0112)	(0.0183)
Sample size	723	360	363	736	306	429
Panel B: Gender	Sons			Daughters		
Maternal lifespan	0.1461***	0.2049***	0.0868*	0.2387***	0.1683**	0.2820***
	(0.0403)	(0.0599)	(0.0499)	(0.0531)	(0.0851)	(0.0721)
Maternal lifespan × Income score	0.0228***	0.0128	0.0304**	0.0225	0.0071	0.0341
	(0.0085)	(0.0109)	(0.0139)	(0.0149)	(0.0208)	(0.0232)
Sample size	868	402	465	591	264	327
Panel C: Ages at Death	Including those who died before age 20			Using both children and parents who survived up to age 50		
Maternal lifespan	0.2168***	0.2186***	0.2290***	0.1318***	0.1371**	0.1118**
	(0.0445)	(0.0635)	(0.0607)	(0.0385)	(0.0685)	(0.0477)
Maternal lifespan × Income score	0.0412***	0.0262**	0.0505***	0.0124	0.0077	0.0133
	(0.0110)	(0.0126)	(0.0176)	(0.0101)	(0.0137)	(0.0155)
Sample size	1722	783	934	867	386	480

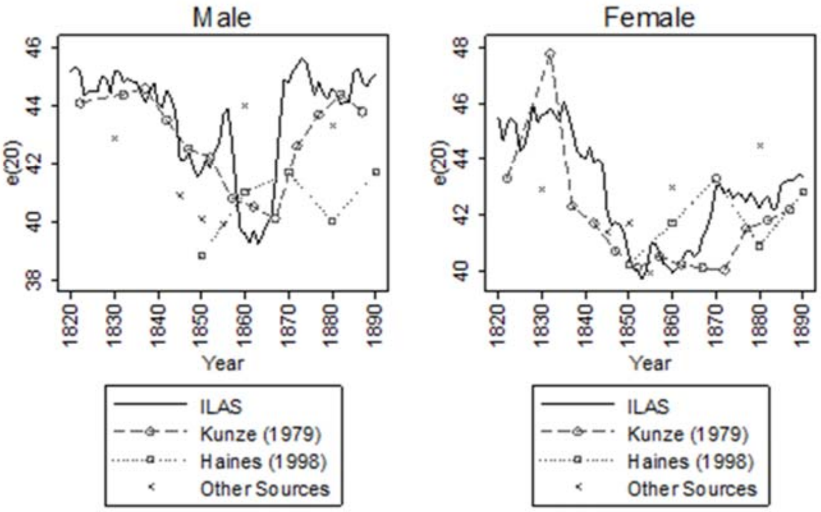
Note: We conduct the same regressions as done in models (1) and (2) of Table 3 for two different groups clustered by birth decade in panel A, children's gender in panel B, and ages at death in panel C. We assume that the group categorized as high risk was exposed to a high risk of premature death in the mid-nineteenth century. In panel A, cohorts born in 1865-1884 and 1845-1864 are high and low-risk group, respectively. In panel B, we use sons for high-risk group and daughters for low-risk group. In panel C, we contain children who died before age 20 in the sample for the high-risk group; the low-risk group includes only children and parents who survived up to age 50. Robust standard errors are reported in parentheses. A single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.

Table A1. Summary of Socioeconomic Status: ILAS vs. IPUMS

Variables	ILAS	IPUMS	p-value
Panel A: Parents cohort - males aged 20-50 in 1870			
Age	35.5	34.0	0.000
Real property wealth	1735.5	1529.7	0.052
Personal property wealth	638.0	721.3	0.812
Total wealth	2373.5	2250.9	0.272
Ratio of farmer	0.2627	0.3164	1.000
Rate of illiteracy	0.0303	0.0691	1.000
Regional distribution			
Northeast	0.3720	0.3559	0.116
Midwest	0.3055	0.3934	1.000
South	0.3163	0.2224	0.000
West	0.0062	0.0284	1.000
Observations	3236	48099	
Panel B: Children cohort - Those aged 0-20 in 1870			
Age	9.1	9.6	1.000
Ratio of boy	0.5133	0.5039	0.057
School attendance rate	0.2561	0.3972	1.000
Observations	7440	157394	

Note : P-value tests whether the difference of sample means is statistically significant or not

Figure A1. Comparison of Life Expectancy at Age 20 from the ILAS Data and Other Sources



Note: e(20) from the ILAS data is 10-year moving average of annually estimated e(20) aournd the given year. Other sources include Elliot(1857) for 1855, Billings (1885) for 1879-82, Meech (1898) for 1830-60, Jaffe and ourie (1942) for 1826-35, Jacobson (1957) for 1849-50, and Vinovskis (1972) for 1859-61. The estimates from other sources are plotted at the center of survey years.

Table B1. Summary Statistics of Individual Level Variables

	All Sample		Children who died before age 50		Children who survived up to 50		Data source
Variables	mean	sd	mean	sd	mean	sd	
Children's age at death	64.47	18.80	35.66	8.41	72.80	11.21	ILAS
Mother's age at death	67.13	16.92	63.24	17.33	68.26	16.63	ILAS
Father's age at death	71.35	13.77	69.30	15.11	71.94	13.31	ILAS
Mother's age at the child's birth	30.30	6.60	30.88	6.67	30.14	6.57	ILAS
Father's age at the child's birth	35.23	7.90	35.76	7.59	35.07	7.99	ILAS
Dummy of male	0.59	0.49	0.53	0.50	0.61	0.49	ILAS
Number of siblings	7.44	3.13	7.41	3.25	7.45	3.10	ILAS
Father's occupational income score at the child's birth	17.81	10.25	17.87	10.28	17.80	10.25	ILAS, IPUMS
Dummy of being born in large cities	0.049	0.217	0.067	0.251	0.044	0.206	ILAS, US Census
Observations	1459		327		1132		