NBER WORKING PAPER SERIES

THE ASSOCIATION BETWEEN EDUCATIONAL ATTAINMENT AND LONGEVITY USING INDIVIDUAL LEVEL DATA FROM THE 1940 CENSUS

Adriana Lleras-Muney
Joseph Price
Dahai Yue

Working Paper 27514
http://www.nber.org/papers/w27514

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020

We are grateful to Sena Ustuner and Evan Rittenhouse for excellent research assistance. This project was supported by the California Center for Population Research at UCLA (CCPR), which receives core support (P2C- HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Association Between Educational Attainment and Longevity using Individual Level Data from the 1940 Census
Adriana Lleras-Muney, Joseph Price, and Dahai Yue
NBER Working Paper No. 27514
July 2020
JEL No. I10,I20,J10

ABSTRACT

We combine newly released individual data from the 1940 full-count census with death records and other information available in family trees to create the largest individual data to date to study the association between years of schooling and age at death. Conditional on surviving to age 35, one additional year of education is associated with roughly 0.4 more years of life for both men and women for cohorts born 1906-1915. This association is close to linear but exhibits strong credentialing effects, particularly for men, and is substantially smaller for cohorts born earlier. This association varies substantially by state of birth, but it is not smaller in states with higher levels of education or longevity. For men the association is stronger in places with greater incomes, higher quality of school, and larger investments in public health. Women also exhibit great heterogeneity in the association, but our measures of the childhood environment do not explain it.

Adriana Lleras-Muney
Department of Economics
9373 Bunche Hall
UCLA
Los Angeles, CA 90095
and NBER
allerasmuney@gmail.com

Joseph Price
Department of Economics
Brigham Young University
162 FOB
Provo, UT 84602
and NBER
joseph_price@byu.edu

Dahai Yue
Department of Health Policy
and Management
650 Charles Young Dr. S., 31-269 CHS
UCLA
Los Angeles, CA 90095
dhyue@ucla.edu

1. Introduction

Educational attainment is a profound predictor of longevity. Prior studies across different disciplines and in various countries show that those with more years of schooling had a lower mortality risk compared to those with less education (Buckles et al. 2016; Hummer and Hernandez 2013; Kitagawa and Hauser 1973; Lleras-Muney 2005). In the US, these educational disparities in mortality have increased dramatically since the mid-1980s across race and gender groups (Case and Deaton 2015, 2017; Hayward et al. 2015; Meara et al. 2008; Montez et al. 2011; Olshansky et al. 2012; Sasson 2016). Understanding the nature of this relationship and the reasons why it is so strong and persistent is a key issue for both researchers and policy makers.

In this paper, we investigate the relationship between education and longevity, using a novel and very large individual-level dataset for the US. We combine recently released individual data from the 1940 full US census with death records and information from family trees in FamilySearch. FamilySearch has a wiki-style family tree with over 1.2 billion people—the largest collection of its kind. We study the association between years of schooling and age at death (also referred to herein as longevity or lifespan) among more than 5 million white individuals born in the US between 1870 and 1915 who were alive in 1940.

The majority of previous US research on the education-mortality relationship used data from the National Vital Statistics System, survey datasets linked to the National Death Index, or census data.[1] However, these datasets have several limitations. First, although the National Vital

[1] Unlike multiple European countries (where previous studies have used large individual data sets to study this question, e.g., Behrman et al. 2011, Lager and Torssander 2012), the US does not have national registry data.

Statistic System is considered the single most comprehensive data source on US mortality, it suffers from well-known measurement errors in educational attainment. Educational attainment available on death certificates is reported by funeral directors and other individuals instead of the deceased. It is thus rounded (heaped) at 12 years of schooling from both lower and higher-level education (Rostron et al. 2010; Sasson 2016). Our data rely on years of education that are reported to the census enumerator while the person is still alive.

Second, other studies have used cross-sectional surveys linked to the National Death Index, such as the National Health Interview Survey Linked Mortality File and the National Longitudinal Mortality Study (Case and Deaton 2017; Hayward et al. 2015; Montez et al. 2012). Educational attainment is better measured in these surveys, but these datasets are typically limited by sample sizes, and can only track mortality within a short period.[2] Because we track the age at death in family trees, we observe deaths occurring from 1940 until today.

We focus on individual-level longevity rather than aggregate mortality risk, within a defined time frame, providing the first estimates of the relationship between longevity and education at the individuals level. We find large associations between education and longevity: one more year of schooling is associated with about half a year longer life for both men and women. We show that estimates of the education gradient are grossly underestimated when the data are censored, for example, when using death certificates obtained during a finite period, such as the Death Mortality Files.

[2] Alternatively, census data has also been used in studies of the education-mortality relationship (e.g., Lleras-Muney 2005). In these studies, mortality suffers from substantial measurement error (Black et al. 2015).

We test the functional form of this relationship and confirm findings in previous work that suggest important credentialing effects. We also confirm previous findings that in either logs or levels, the education gradient in longevity diminishes with age (Crimmins 2005; Elo and Preston 1996; Kitagawa and Hauser 1973; Lynch 2003). We also find that the education gradient is larger for more recently born cohorts (Meara et al. 2008; Montez et al. 2011; Montez et al. 2019; Olshansky et al. 2012).

Taking advantage of the large size of our sample, we investigate whether education gradients vary geographically and why. Recent work suggests that place of residence is a strong predictor of mortality (Chetty et al. 2016; Deryugina and Molitor 2018; Finkelstein et al. 2019). Moreover, Montez et al. (2019a,b) document that the association between education and mortality rates differs by state-of-residence. We focus on state-of-birth instead to overcome the potential issue of selective migration: education might affect the decision to migrate and where to live (Currie and Schwandt 2016). But individuals do not choose their state-of-birth.

There is substantial variation in the association between education and longevity based on state-of-birth. This variation across states is similar across genders but there are some important differences. For men the education gradient in longevity is *larger* in places with greater education and greater baseline longevity, whereas for women the gradient does not vary by these baseline levels. These results do not support a simple view of decreasing returns to scale—we can rule out that the associations fall with baseline education or longevity levels.

We correlate these state-level education gradients with various state-level measures to explore if childhood circumstances modify education gradients, which previous work suggests have long-lasting economic and health effects (Almond et al. 2018). We explore three theories. First, we hypothesize that the returns to a year of schooling are larger in places where the quality of education is higher. To test this, we focus on the measures of quality of schooling compiled by Card and Kruger (1992), who first documented that the association between education and wages was larger for men who went to school in places with greater quality of schooling. We find that the association between education and longevity is also larger for men who went to school in places where the quality of school was higher. But there's no such relationship for women.

Next we explore whether the education-longevity association varies depending on the level of public health investments. On the one hand, it's possible that when mortality levels are high, and there are few health resources that can affect health, education will matter less. In this case, we might expect that the association between education and longevity will increase as mortality levels fall, because more educated individuals will be more likely to use the knowledge and technologies that cause mortality to fall. On the other hand, public health interventions like sewers and water filtration, which lowered infant mortality in the early 1900s (Alsan and Goldin 2019; Cutler and Miller 2005), often benefit all, irrespective of education. In this case education gradients might fall as health levels increase, because such interventions disproportionately benefit groups with high initial mortality levels.[3] Finally it is also possible that child health and education are complements (or substitutes) in the production of adult health and mortality. We

[3] For example, Acemoglu and Johnson (2007) show that innovations like DDT and penicillin increased longevity more in places with an initially large incidence of malaria and infectious disease that antibiotics treat. Becker, Philipson, and Soares (2005) show these innovations lowered inequality in longevity across countries of the world.

find that education gradients are larger in places with more health resources for men, but not clearly for women.

Lastly, we assess whether the education-longevity association is larger in places with greater incomes. It's long been hypothesized that part of the association between education and longevity (or other health measures) is due to the greater financial resources during childhood that allow individuals to gain high levels of education and better health status. If this hypothesis is correct, the education gradient in longevity and childhood financial resources should be positively associated. It is also possible that education and family income are complements in the production of adult health. For example, better fed children might benefit more from education. We find that the education gradients are larger for those born in states with higher per capita income for men but not for women.

2. **Data and Descriptive Statistics**

2.1. Data

Primary micro-level data. The primary microdata comes from the 1940 full-count US census—the first national census to collect information on years of schooling and income for almost all respondents, including name, place of birth and year of birth for all respondents.

These data were matched to the FamilySearch database, a genealogy platform with over 12.6 million registered users, and profiles for over 1.2 billion deceased individuals. When people search their own family histories, they gather source documents, including census records, and upload information (e.g., vital events) to the profiles of their ancestors on a genealogical website like FamilySearch. These data include name, date of birth, date of death, dates of marriages and links between parents and children.

Our outcome of interest is the age at death (or *longevity*—we use these terms interchangeably), which is measured in years and computed as the death year minus the birth year.[4] The main explanatory variable of interest is years of education, which ranges from zero to 17, corresponding to five or more years of post-secondary education.

Sample selection. We exclude those born in Alaska and Hawaii (which weren't yet states), the District of Columbia, or outside the US. We drop individuals missing information on education, date of birth or date of death. We keep individuals who were 25 to 70 years old in 1940 (born 1870 to 1915), who by then had completed their education and were well represented. Because blacks and individuals of other races are poorly represented in family trees, we concentrate on individuals self-reporting as white in 1940.[5] Finally we drop individuals whose age at death appears to be implausible. **Figure A1** shows the details of how we move from the original 1940 census data to the final data, which includes 5.4 million individuals.

For our analysis of the geographic moderators, we focus only on the youngest cohort born 1906 to 1915, for whom we have data on the childhood conditions in their state-of-birth. In order to

[4] Month and day of birth and death are missing for a substantial number of cases, so we rely on year of birth and year of date only.

[5] Although about 10 percent of the population in the 1940 census is black, in the matched data they only account for 0.6 percent of the observations. Fully understanding the under-representation of black individuals in our data and its implications for the education gradient is beyond the scope of this paper.

make them comparable, we restrict attention further to those who are alive at age 35. This sample has about 1.36 million observations. Another reason to focus on this young cohort is that the education of older cohorts is overstated in the 1940 census (Goldin 1998).

State-level data for the 1906-1915 cohort

*Quality of Schooling by state***.** Based on issues of the *Biennial Survey of Education*, which contains the results of surveys conducted by the US Office of Education from 1918 to 1966, Card and Krueger (1992) compiled a dataset measuring the quality of public schools, based on the ratio of enrolled students to instructional staff in the state (pupil/teacher ratio), the average length of the school term (term length), and average annual teacher salaries. Stephens and Yang (2014) extended the data series to birth cohorts from 1905 to 1959, using various editions of the *Digest of Educational Statistics*.[6] We use the average of quality measures for the 1906-1915 cohorts.

*Child mortality, number of doctors and number of nurses from the 1910 census***.** We construct a measure of child mortality using the 1910 census, which asked women the number of children they ever had, and the number of children they had that died. We use the fraction of children that died to women ages 16 to 45 as a proxy for child mortality. This measure ranges from 129 in Iowa to 294 in New Mexico, with an average of 185 deaths per thousand.[7] Using the occupation questions in the census, we compute the total number of doctors and nurses in each state, divided

[6] The dataset is available on the journal website: https://www.aeaweb.org/articles?id=10.1257/aer.104.6.1777

[7] Official infant mortality rates by state are only available for a few states with complete birth and death registration systems. We collected data for these states. The correlation between our measure and the 1900 infant mortality for the 8 states (CT, MA, ME, MI, NH, NY, RI, and VT) for which official measures are available is 0.97. The correlation in 1915 is 0.92, based on official measures from 10 states (CT, MA, ME, MI, MN, NH, NY, PA, RI, and VT).

by the state population. The number of doctors per thousand ranges from 0.69 in South Carolina to 2.15 in Colorado, averaging 1.35 across states. The number of nurses per thousand ranges from 0.07 in Oklahoma to 0.65 in California, averaging 0.28 across states.

Per capita income by state. We use estimates of state-level per capita income (in 1929 dollars) in 1900 and 1920, reported by the Bureau of Labor Statistics.[8] It ranges from $211 in Mississippi to $818 in Nevada, averaging of $478.4 across states.

2.2 Summary Statistics, Representativeness and Data Quality

Summary statistics for the data are in **Table 1**. The average years of schooling is nine for the full sample, 10 for the most recent cohort, and is slightly larger for women. These averages conceal much variation. **Figure A2** shows the distribution of education by gender for the 1906-1915 cohort. Although more men graduate college, many more women graduate high school. The larger education of females for these cohorts has been documented elsewhere: women had higher high school graduation rates in every state for every year from roughly 1910 into the 1930s (Goldin 1998).

For the birth cohorts born between 1870 and 1915, the mean age at death is 76.2, but is five years greater for females (79.1) than for males (73.9). For the most recent cohort, conditional on surviving to age 35, the gender gap in longevity is six years. The increase in the gender gap is greater if we condition on being alive at age 65 for all cohorts, consistent with the growing survival advantage of women documented elsewhere (Barford et al. 2006; Beltrán-Sánchez et al.

[8] These data come from estimates produced by Kuznets, Miller and Easterlin (1960).

2015; Cullen et al. 2016; Goldin and Lleras-Muney 2019; Preston and Wang 2006).[9] **Figure A3** shows the entire distribution of the age at death by gender for the 1906-1915 cohort. The distribution of longevity is not quite normal—there is a long left tail of early deaths, particularly for males. Notably, there is no evidence in the data that there is age heaping, suggesting that the data on dates is of high quality.

Representativeness. Figure A2 compares the distribution of education in the 1940 census and in our matched sample, for whites born in the contiguous 48 states from 1906 to 1915. The distributions are very similar, with few small differences: we have more individuals with exactly 8 or 12 years of schooling, whereas there are more individuals with college degrees in the full 1940 census. But the differences are small. We cannot reject the null that the distributions are identical using the Kolmogorov–Smirnov test. (For men the p-value is 1; for women it's 0.964.)

Table A1 shows that the census-tree linked data are also representative of education for each birth cohort born between 1906 and 1915. The sample represents roughly 10% of the original census count. The table also shows that the census-tree linked data contains more males compared to the 1940 full-count data. In general, it is more difficult to trace women through historical records because their surname changes. However, our ability to track women is much better than in other historical research.

Table A2 shows the distribution of the observations by state-of-birth for the youngest cohort. The spatial distribution in our data differs from the distribution in the census in some important ways. Individuals from the Midwest are over-represented in our data, whereas individuals from

[9] The gap in average longevity between men and women at age 65 rises from 2.53 for the 1876-1885 cohort to 4.48 for the 1906-1915 cohort.

the Northeast are underrepresented. Most notably, the most populous states (CA, NY, NJ, PA, TX) are underrepresented in our data, with the exception of Ohio.[10] These differences however are not statistically significant—we cannot reject that the distributions are equal (p-value for male: 0.368; for female: 0.687). We assess the sensitivity of our results to weighting schemes to make the data nationally representative.

Quality of death information. An important consideration is whether the age at death information in the family tree, which comes from multiple sources, is of high quality. One way to assess this is to compare our age at death to the expected age at death reported in the Social Security Administration (SSA) cohort life tables. These tables show that for the 1910 cohort, conditional on surviving to age 35, average age at death is 71.61 for males and 78.54 for females.[11] In our data, the average age at death for this birth cohort is 72.69 for males and 79.62 for females, about a year higher than in the SSA data. This difference is to be expected because blacks, immigrants and low SES individuals are less likely to be represented in our data, but are included in SSA computations.[12]

We also check the quality of our data against newly released version of the 1940 census matched to mortality records created by a research group at Berkeley, the CenSoc-Numident.[13] This database matches all individuals observed in the 1940 census to death certificates of individuals who died between 1988 and 2005 in the Numident file held by the SSA. The match is done by

[10] This is partly due to the fact that historical records for these states are poor, whereas they are excellent in Ohio and other midwestern states like Illinois, Indiana, Iowa, and Idaho. It is also likely due to the fact that second generation immigrants are less likely to be in the family trees as are individuals of low SES.
[11]These tables are available here: https://www.ssa.gov/OACT/NOTES/as116/as116_Tbl_7_1910.html#wp1081274
[12] There could be other sources for the differences. The SSA computations are based only on data from states with registration systems. The SSA attempts to make these data representative by weighting the state level data. Using weights does not materially change our averages (weighted longevity is 79.45 for females and 72.63 for males).
[13] These data are available here: https://censoc.berkeley.edu/

first name, last name, year of birth, and place of birth. There are several differences between our data and this database. First, our data contain information on the age at death from multiple sources. As a result, our database includes deaths across all possible years, not just those occurring 1988 to 2005. Second, our data is constructed differently, relying more on people, who generated most of our matches to death records—not on algorithms alone. It is not clear that our data is more accurate for any given person or group. On average, however, we expect similar results.

Figure A4 shows the average age at death by birth cohort in the CenSoc-Numident and in our data. To make the samples comparable in both datasets, we restrict attention to those born in the contiguous 48 states categorized as white in the 1940 census, who died between 1988 and 2005. The mean age at death in both datasets is very similar (differences are less than 0.5 years in all cases), exhibiting a similar (downward) trend for men and women.[14]

The data differ in other dimensions. **Table A3** shows how the two datasets compare. Longevity is greater in our data, despite similarities in the preceding figure, because of differences in the extent to which different cohorts are represented. (The CenSoc-Numident data include a larger share of older cohorts.) The geographic distribution of the CenSoc data is also different with more observations from the Northeast and fewer from the Midwest. We assess the extent to which our results differ from those derived from this alternative data throughout the paper.

[14] The age at death is falling on both data because they both condition on being alive in 1940. So although more recent cohorts live longer, older cohorts are only observed in 1940 if they lived long enough to be alive in 1940. This selection causes the downward trend observed in both data sets.

3. Basic Associations Between Education and Longevity

Panel A Figure 1 shows some preliminary evidence of the association between education and longevity in our data for the 1906-1915 cohort. It shows that the average age at death increases with education almost linearly for both genders. **Panel B** further shows the estimated density of the age at death conditional on surviving to age 35, for various education groups: no school (0), some elementary (1-7), some high school (8-11), some college (12-15), and college plus (16+). For both men and women, the density of longevity shifts right when education increases.

To estimate the extent of the association, we estimate the following regression model:

$$y_{ics} = \beta_0 + \beta_1 \times education_{ics} + \boldsymbol{X}_{ics}\theta + \mu_s + \gamma_c + \delta_{sc} + \varepsilon_{ics} \qquad (1)$$

where the outcome *y* is the age at death for individual *i* born in year *c* in state *s*. Education is the number of years of schooling in the 1940 census for the individual. The regression also includes individual characteristics such as gender ($\boldsymbol{X}_{ics}$), dummies for each birth cohort (γ_c), for each state-of-birth (μ_s) and state-of-birth specific cohort trends (δ_{sc}). $\beta_1$is the coefficient of interest, measuring the association between education and longevity, which we also refer to as the education gradient in longevity. We report Huber-White robust standard errors clustered at the state-of-birth level.

Table 2 shows the results for the 1906-1915 birth cohorts conditional on being alive at age 35. Column 1 includes no covariates. Column 2 controls for cohort fixed effects. Column 3 controls for state-of-birth fixed effects and column 4 adds state-specific linear trends. The coefficient of education is positive and statistically significant in all regressions and for both genders. The estimates are remarkably stable across columns.

Conditional on surviving to age 35, one additional year of education is associated with roughly 0.4 more years of life. Women benefited slightly more (0.43) from the education than men (0.40)—but this difference is not statistically significant at the 5% level (p-value=0.193). Relative to mean longevity (72.79 for men; 79.63 for women), this effect is roughly equivalent to a 0.5 percent increase in longevity for each additional year of schooling. Alternatively, an increase of one standard deviation in education would increase longevity by 1.2 years for both men and women. The estimates are very similar if we weight the observations to make them representative of the nation (column 5). Once we do this, there is no longer any difference between genders.

We also estimate an Accelerated Failure Time model (AFT), common in demography, which uses the log of the age at death as the dependent variable. The results from this estimation (in **Table A4**) lead to similar conclusions, with one more year of education increasing longevity by about 0.6 percent. We prefer the model in levels because the distribution of the age at death has a fat left tail, which is less consistent with the log normal assumption.[15]

Data quality and effects of truncation. We now compare our results to the results one would obtain using Berkeley's CenSoc-Numident data. **Table 3** shows the results. The education gradient in the CenSoc-Numident data is much smaller (0.088) than in our data (0.41). Columns

[15] In fact, the data reject both normality and log normality for the distribution of the age at death.

2 and 3 show this is the result of limiting the years during which deaths are observed. In column 2, we show that if we restrict our data to include only individuals who died between 1988 and 2005 then our estimates are close, where the coefficient of education is 0.104 compared to 0.088 in the CenSoc-Numident data. This suggests that the matching approach does not affect the results.

These results show that our data is of similar quality to the CenSoc-Numident data. They also highlight the benefit of using the census tree data. In column 3 we show that when we include all deaths from 1941 to 2005, we get a dramatic increase in the coefficient of education, which rises from 0.10 to 0.35. Thus, including early deaths makes a very large difference to our estimates. Column 4 shows that if we include deaths up to 2019, the coefficient of education rises to 0.42, another substantial (but smaller) increase.[16] These results document that left and right censoring effect a substantial attenuation in the estimates of the education gradient, because more educated individuals are more likely to survive to 1988, and to live beyond 2005.

Estimates by cohort and age. So far, our results only show associations for the youngest cohort in the data. **Figure 2** shows the associations by cohort and survival age. For each ten-year birth cohort, we restrict the sample so that everyone has survived to the same age: the 1906-1915 cohort is restricted to surviving to age 35, the 1896-1905 cohort is restricted to survive to age 45, etc. **Panel A of Figure 2** plots the results when gradients are estimated in levels.

[16] There are some small differences by gender. The education gradients for men are very close on both data sets (0.099 in the CenSoc data vs 0.108 in the census-tree data). For women the CenSoc data produces lower estimates 0.0788 compared to 0.0992.

We find that the relationship between education and longevity is greater for more recent cohorts than for older cohorts surviving to the same age. For example, conditional on surviving to age 55, one more year of education is associated with 0.34 years of life for the 1906-1915 cohort, but only with 0.25 years of life for the 1896 cohort and 0.17 years of life for the 1886-1895 cohort. The gradients are about twice as large for the more recent cohort than for cohorts born 20 years earlier. More recent cohorts are also more highly educated—thus the gradient is increasing with the level of education across cohorts.[17]

In the 1940 census educational attainment, particularly high school graduation, is overstated for cohorts older than 35 (those born before 1905; Goldin 1998); and so caution is needed in interpreting these results. If individuals with low levels of schooling, who lived short lives, reported higher levels of schooling, the relationship between years of school and longevity would likely be attenuated, since the average longevity of the highly educated group would fall as a result of the misclassification.[18] Goldin finds that these errors are smaller in states with higher educational attainment. In **Fig. A5** we show that the results are very similar among high education states, suggesting that the increase in the education gradient across cohorts is real rather than due to measurement error for older cohorts.

We also find that the education gradient is lower at older ages. This is true for all cohorts. For example, with the 1906-1915 cohort, conditional on surviving to age 35, one more year of

[17] In the 1940 census, the mean education was 8.26 for the 1876-1885 cohort, 8.72 for the 1886-1895 cohort, 9.29 for the 1896-1905 cohort and 10.04 for the 1906-1915 cohort. These differences in educational attainment across cohorts are likely to be underestimated since those with lower education are less likely to have survived to 1940.
[18] The effect of misclassification on the education gradient is ex-ante unclear. If low education individuals who live long lives report higher education, then this would result in an over-estimation of the results.

education is associated with 0.41 additional years of life, but only with 0.28 years of life for those surviving to age 65. This evidence is consistent with the idea that education plays an important role in preventing early deaths. Analyses that condition on survival to old age will find lower estimates of the gradient. This result confirms observations in the literature based on alternative data and estimation methods, e.g. Kitagawa and Hauser (1973), Elo and Preston (1996), Lynch (2003), and Crimmins (2005). These conclusions are identical if we estimate a linear or a log linear model as shown in **Panel B of Fig. 2**.

4. Heterogeneity by state of birth for the 1906-1915 cohort.

Previous work has documented substantial heterogeneity in the association between education and various health measures, including mortality by gender (Montez et al. 2011; Ross et al. 2012), race (Montez et al. 2011; Williams and Jackson 2005), overtime (Bound et al. 2015; Cutler et al. 2015; Goldring et al. 2016; Meara et al. 2008; Olshansky et al. 2012), and state-of-residence (Montez et al. 2019; Montez et al. 2019). In this section, we focus on spatial variation by state-of-birth. Because we only have state-level data for the most recent cohort (1906-1915), this analysis focuses on them.

We estimate education gradients by state of birth and gender, controlling for birth cohort dummies as in Eq. 1 above. The estimated coefficients on education are positive and statistically significant for both men and women in all states. However, there is wide variation (**Fig. A6**). **Table A5** shows the top ten and bottom ten states ranked based on the education gradient. In Utah (the top state), one more year of education is associated with 0.72 additional years of life, whereas in New Mexico (the bottom state) one more year of education is associated with only

0.25 more years of life. Moving from the bottom 10th percentile of the education distribution to the 90th yields 4.3 more years of life in Utah but only 2.3 in New Mexico. These large differences are statistically significant at the 5% level, and similar if we estimate gradients based on a log model (Panel B). Utah remains the top state even when we account for the fact that life expectancy in Utah is long. Similarly, New Mexico and South Dakota are the two bottom states in either levels or percentages.

Figure 3 documents this variation in space by gender. For men education gradients were generally larger in the West and Northeast, as well as in Indiana and Ohio. The associations are smaller in the South, except for Florida and Louisiana, and in many states in the middle of the country. In general, states with large associations for men also have large associations for women (**Fig. A7**). But there are noticeable exceptions. In California, Louisiana and Florida the associations between education and longevity are very small for women but above average for men. Conversely in Tennessee, Kentucky and Missouri, the association is large for women, but small for men. Overall and quite surprisingly the correlation in the male and female education gradient across states is small (0.26). Thus, while the overall gradient is not different by gender, there appear to be large differences in why/when education is associated with longevity, differing by gender and space.

In **Fig. A8**, we explore whether these gradients are stable over time. To do this, we re-estimate the gradient conditioning on survival to age 65 for all cohorts. Then we correlate the gradient for the 1904-1915 cohort with the gradient for the oldest 1876-1885 cohort. Surprisingly, the gradients are not very related. The relationship is positive and statistically significant for males (regression coefficient = 0.286, p = 0.012) but negative and not significant for females (regression coefficient = -0.115, p = 0.227). This suggests that there were significant changes throughout the period affecting the association between education and longevity. These changes led to greater associations for more recent cohorts and to different associations based on state of birth. This suggests that environmental characteristics, which were changing throughout the period, resulted in vastly different education gradients.

We now explore reasons for this heterogeneity.

One possible explanation is that the association between education and longevity exhibits decreasing returns to scale, with greater "returns" at low levels of education. To test this, **Fig. A9 panel A** plots the gradients by state-of-birth against the average level of education in the state for the same cohorts. Contrary to expectations, for men, the education gradients are larger in places where education was already great to begin with. A regression estimates that for men, one more year of education increases the education gradient by 0.11, a very large increase relative to the mean association of 0.5. By contrast, for women, the education gradient is flat and does not vary with the level of education.[19] Thus for neither men nor women do we find that the gradient is smaller when education is greater.

Panel B investigates a related hypothesis: that the education gradient decreases when longevity increases. Again we find that for men the education gradient is greater in places with greater

[19] The slope of the regression of the education gradient on years of education is 0.1072 ($p < 0.001$) for men and 0.0118 (p=0.528) for women. We can reject the relationship is negative for men ($p < 0.001$) but not for women (p = 0.264)

longevity, though this association is not statistically significant. For women, we find a small decrease, but the decrease is not statistically significant either. These results are very similar if we estimate state- and cohort-specific gradients and plot them against education or longevity levels (**Fig. A10**). Thus, the evidence does not support the hypothesis that there are decreasing returns based on education or longevity levels.

Next, we explore whether the returns to education vary with the quality of education. To test this, we make use of state-level data containing quality of schooling measures. Note that education levels and quality of education are only mildly correlated –there is variation in the quality of schooling within the same level of education.[20] **Figure 4** plots the estimated education gradients by gender against our measures of quality: relative teacher wages, length of school term and pupil-teacher ratios. For males all three measures strongly predict education gradients (Panel A). Put otherwise, when teachers were well paid, the school year was long, and pupil-teacher ratios were low, men who went to school longer benefited more from schooling in terms of their longevity. The results for men are consistent with the findings of Frisvold and Golderstein (2011) who show, using the same measures of quality of school, that higher quality of education led to greater health gradients among Blacks born between 1930 and 1950 in the South.

However, Panel B shows that the same does not hold true for women. Quality of schooling measures exhibit only a weak association with the education gradient. **Table A6** shows these results quantitatively. If we regress education gradients on school measures, individually or jointly, the regressors are statistically significant predictors for men, but not for women.

[20] The correlation between years of education and relative teacher wages is 0.18, 0.22 for term length and -0.19 for pupil teacher ratios.

The second hypothesis is that the education gradient varies with the level of health resources. We use three measures observed in 1910: child mortality, number of doctors and number of nurses. **Figure 5** shows the results. For men (Panel A), we find that the education gradient is larger in places with better resources (lower child mortality, higher number of doctors or nurses). But again these associations are much more muted for women. **Panel B of Table 6** confirms these findings. Male gradients are statistically significantly correlated with all three measures, most notably with the number of nurses. These findings support the hypothesis that innovation increases health inequality consistent with the "fundamental causes of disease" hypothesis by Link ant Phelan (1995) and Link et al. (1998) and documented by Glied and Lleras-Muney (2008). But only the number of doctors predicts education gradients for women, and jointly we cannot reject that all three measures do not matter for women's gradients.

The last hypothesis we explore is that health associations are larger in richer states, measured by per capita income. **Figure 6** shows that this is true for men but not for women. **Panel C of Table A6** again confirms the associations are statistically significant for men but not women.

In sum, for men, we find that the large variation in education gradients across state-of-birth can be explained by differences across states in the quality of schooling, the level of public health resources and state-level incomes. However, this section ends with a puzzle. Although education is equally predictive of longevity for women and men, the spatial variation in the education gradient is quite different by gender, and we found no early life indicators explaining this variation for women. Indeed, if we regress the education gradients at the state level on all state characteristics (**Table A7**), we find that the r-squared in this regression is high for men (0.57), and much lower for women (0.18). We return to this question in the discussion section.

5. Testing for linear relationships

So far, we have estimated linear relationships between education and longevity (or its log). However, a large literature has argued that there are important credentialing effects. Because in our data we observe education as a continuous measure ranging from 0 to 17, we can estimate strict non-parametric models of the education-longevity relationship, and test whether this relationship is linear or subject to credentialing (or sheepskin) effects. A close look at **Fig. 1 panel A** suggests that while the relationship looks linear, there are visible jumps at 1, 8, 12 and 16 years of education, though these jumps are much more visible for men than for women.

To investigate this, for the youngest cohort, we estimate a fully non-parametric model, where 0 is the left-out category. We include a dummy for every single year of school. The models control for state-of-birth dummies, year-of-birth dummies, and state-of-birth-specific linear trends, and are estimated separately by gender. The estimated coefficients are plotted in **Fig. 7 Panel A.** (The point estimates are in **Table A8.**) These estimates show that for men the relationship between education and longevity is best described as a series of step functions with increases at 1, 8, 12 and 16, and no increases in between. For women the relationship is more linear before 8 years of school, but becomes a step function thereafter. Results from the CenSoc-Numident are similar for both men and women (**Fig. A11**). **Table A9** shows that we reject the linear specification for both men and women for the full range, but not for education levels between 1 and 6.

In **Table A10** we present the fit of different models: a fully non-parametric model, a linear model, and a model with splines at exactly 8, 12 and 16 years of school, which correspond to concrete degrees. We show four measures of fit: the adjusted r-square, the AIC, the BIC and the mean squared error from a cross-validation exercise.[21] The spline model provides the best fit (highest adjusted r-square, lowest AIC, BIC or MSE) for both men and women. However, the linear model still provides an excellent fit: our fit measures do not improve much by moving from the linear to the spline or the non-parametric model.

Figure 7 Panel B shows the evolution of the gradient for different cohorts that results from estimating the non-parametric model for various cohorts, restricting all cohorts to survival to age 65 so the numbers are comparable across cohorts. Several conclusions emerge. First there is a very large difference in the pattern between men and women. The associations for women are quite close to linear for almost all cohorts. For men credentialing effects are much more important. For men in more recent cohorts the increases in longevity associated with 8, 12 and 16 years of school are larger. This results in a fanning out of the relationship: there are very few differences across cohorts for low levels for schooling but an increasingly large difference at higher levels of schooling across cohorts. For women, instead, we observe that there are larger associations for more recent cohorts for almost all levels of schooling. Overall the gradients are larger for more recent cohorts consistent with our earlier findings.[22]

---

[21] We use a 10-fold cross-validation process to compute the MSE. Specifically, we first randomly shuffle the dataset and split it into 10 groups. For each unique group, we take the group as a holdout, estimate the model on the remaining groups, compute the MSE based on all groups, and store the mean MSE. We then take the average of these 10 mean MSE.

[22] **Figure A12** shows these conclusions hold if we estimate the gradients for high education states to minimize the effect of measurement error for older cohorts.

6. Discussion

This paper uses a new individual-level dataset to study the association between years of schooling and longevity in the US across white cohorts born between 1870 and 1915. We find that education gradients are large, grew across cohorts, and get smaller as individuals age. We also find that the relationship is close to linear, particularly for women, though there are important credentialing effects, particularly for men.

This evidence on the functional form of the education-longevity relationship is mostly consistent with existing evidence. Previous studies have depict a linear decline in mortality risk as education increases from 0 to 11, a step-change reduction upon attainment of a high school diploma, and then another steeper linear decline with more years of education (Backlund et al. 1999; Everett et al. 2013; Hayward et al. 2015; Montez et al. 2012). However, we have several new findings. First, for older cohorts, the relationship is almost flat for men whereas it is graded for women in all the birth cohorts we examine. Second, credentialing matters more for men. Third, the "returns" to high levels of education (college) rose quite substantially across cohorts, again despite the fact that more people are attending college.

We document extensive heterogeneity in this association by gender, cohort and state-of-birth. As Hayward et al. (2015) note "…there is no inherent causal association between educational attainment and adult mortality; instead, the causal association is dependent upon time, place, and social environment under study." Our findings are very much in line with this observation.

We observe there is a substantial increase in the association across cohorts, particularly at the upper end of the education distribution. This evidence is consistent with the increase in the education gradients in mortality documented later in the 20th century (Crimmins and Saito 2001; Feldman et al. 1989; Lauderdale 2001; Meara et al. 2008; Montez et al. 2011; Montez et al. 2019; Olshansky et al. 2012; Pappas et al. 1993; Preston and Elo 1995). Many recent analyses of these trends rely on comparisons over time across fixed education categories. As a result many have debated whether these trends are real, or if they reflect changes in the composition of the population within each education category (Bosworth 2018; Bound et al. 2015; Case and Deaton 2017; Cutler et al. 2011; Dowd and Hamoudi 2014; Goldring et al. 2016; Novosad et al. 2020). Our results show that one more year of schooling has a larger return for more recent cohorts, who have greater mean schooling levels, suggesting these increases in disparities are real rather than due to composition effects.

We investigate the sources of heterogeneity across space and show that for men the variation in the association across states is related to environmental conditions in their state-of-birth. But the same is not true for women. In fact, this paper documents that the dynamics of the gradient are substantially different for men and women. While the association is very large for both genders, the association for men can be much more easily explained by variation in their childhood circumstances. For women we also observe substantial heterogeneity, but it is not so easily explained.

Why is the education gradient in longevity moderated by environmental conditions for men but not women? We hypothesize that women are less sensitive to conditions while growing up and that this might explain why gradients are unaffected by these conditions. It has long been hypothesized that males are less biologically buffered than females against the environment during growth and development (Stinson 1985). A few recent studies however find some support for this hypothesis. For example, Doblhammer et al. (2013) find that the effects of being born during the 1866-68 Finnish famine on longevity were large for men but not noticeable for women. Similarly, Lindeboom et al. (2010) find that children exposed in utero to the Dutch Potato famine of 1846-47 had lower longevity, with much larger effects for men. Van den Berg et al. (2016) find that undernutrition between conception and age 4 lowers heights among adult men, but not adult women. Bertrand et al. (2013) find that boys do particularly poorly in disadvantaged environments, and they appear to be more sensitive to inputs. Consistent with these findings, Autor et al. (2019) find that family disadvantage appears to affect boys' educational attainment more than girls'. However, not all research finds differences. For example, Magnusni et al (2016) find no statistically significant differences in the effects of early childhood education interventions, though the point estimates favor larger effects among girls. This male-sensitivity hypothesis related to the large differences in longevity by gender. While environmental forces play an important role in explaining gender gaps in longevity, it is well documented that there are biological differences that disadvantage males in many domains (Kraemer 2000).

Even if the male-sensitivity hypothesis is borne out, it still provides no insight on why there is variation in the education gradients for women. We hypothesize that for women factors like marriage markets and fertility play a large role in moderating the relationship between education and mortality. For example, until the mid 1930s maternal mortality was very large in the US. If more educated women had fewer births, and had them at different ages, then this would result in an education gradient in mortality, both directly through maternal mortality and indirectly since births can have long-term health consequences (like fistulae) that affect the longevity of women who survive past reproductive ages.

Lastly it is possible that gender differences are due to differences in the quality of the data by gender. If we regress our estimates of the education gradients on the fraction of the population that is represented in our data (shown in **Table A2**) we find that the gradients for women are larger in states for which our match is better (beta = 0.0154, p value: 0.026) but not for men. We also find whether individuals migrate and where they migrate is a function of gender, education and its interaction.[23] Women are in fact more likely to move, and thus our measures of the environment they grow up in might be mismatched. Thus, sample selection and differential migration might play an important role in explaining gender differences as well. Understanding these gender differences is an important direction for future research.

This paper has a few limitations. Ideally one would construct cohort-specific measures of childhood conditions—we have done so as much as possible. Data on nutrition and pollution levels would also be valuable—we have only investigated heterogeneity along a few dimensions. Further, we investigate heterogeneity based on state-of-birth. It would be helpful to investigate how many people grow up in the state where they are born; as would replicating the analysis of Montez et al (2011), which compared the effects of state-of-birth and state-of-residence. Finally, we note that we have not documented any causal relationship between education and longevity.

[23] Results available upon request.

Our analysis is strictly descriptive. Future research could exploit exogenous variation in education that comes from compulsory school laws or new college openings.

References

Acemoglu, D., & Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of Political Economy, 115*(6), 925-985.

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature, 56*(4), 1360-1446.

Alsan, M., & Goldin, C. (2019). Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920. *Journal of Political Economy, 127*(2), 586-638.

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2019). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. *American Economic Journal: Applied Economics, 11*(3), 338-381. doi:10.1257/app.20170571

Backlund, E., Sorlie, P. D., & Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: the National Longitudinal Mortality Study. *Social Science & Medicine, 49*(10), 1373-1384.

Barford, A., Dorling, D., Smith, G. D., & Shaw, M. (2006). Life expectancy: women now on top everywhere. *BMJ*, 332:808.

Becker, G. S., Philipson, T. J., & Soares, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *American Economic Review, 95*(1), 277-291.

Behrman, J. R., Kohler, H.-P., Jensen, V. M., Pedersen, D., Petersen, I., Bingley, P., & Christensen, K. (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography, 48*(4), 1347-1375.

Beltrán-Sánchez, H., Finch, C. E., & Crimmins, E. M. (2015). Twentieth century surge of excess adult male mortality. *Proceedings of the National Academy of Sciences, 112*(29), 8993-8998.

Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics, 5*(1), 32-64.

Black, D. A., Hsu, Y.-C., & Taylor, L. J. (2015). The effect of early-life education on later-life mortality. *Journal of Health Economics, 44*, 1-9.

Bosworth, B. (2018). Increasing disparities in mortality by socioeconomic status. *Annual Review of Public Health, 39*, 237-251.

Bound, J., Geronimus, A. T., Rodriguez, J. M., & Waidmann, T. A. (2015). Measuring recent apparent declines in longevity: the role of increasing educational attainment. *Health Affairs, 34*(12), 2167-2173.

Buckles, K., Hagemann, A., Malamud, O., Morrill, M., & Wozniak, A. (2016). The effect of college education on mortality. *Journal of Health Economics, 50*, 99-114.

Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy, 100*(1), 1-40.

Case, A., & Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences, 112*(49), 15078-15083.

Case, A., & Deaton, A. (2017). Mortality and morbidity in the 21st century. *Brookings Papers on Economic Activity, 2017*(1), 397-476.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., . . . Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA, 315*(16), 1750-1766.

Crimmins, E. M. (2005). Socioeconomic differentials in mortality and health at the older ages. *Genus*, 163-176.

Crimmins, E. M., & Saito, Y. (2001). Trends in healthy life expectancy in the United States, 1970–1990: gender, racial, and educational differences. *Social Science & Medicine, 52*(11), 1629-1641.

Cullen, M. R., Baiocchi, M., Eggleston, K., Loftus, P., & Fuchs, V. (2016). The weaker sex? Vulnerable men and women's resilience to socio-economic disadvantage. *SSM-Population Health, 2*, 512-524.

Currie, J., & Schwandt, H. (2016). Mortality inequality: The good news from a county-level approach. *Journal of Economic Perspectives, 30*(2), 29-52.

Cutler, D., & Miller, G. (2005). The role of public health improvements in health advances: the twentieth-century United States. *Demography, 42*(1), 1-22.

Cutler, D. M., Huang, W., & Lleras-Muney, A. (2015). When does education matter? The protective effect of education for cohorts graduating in bad times. *Social Science & Medicine, 127*, 63-73.

Cutler, D. M., Lange, F., Meara, E., Richards-Shubik, S., & Ruhm, C. J. (2011). Rising educational gradients in mortality: the role of behavioral risk factors. *Journal of Health Economics, 30*(6), 1174-1187.

Deryugina, T., & Molitor, D. (2018). *Does when you die depend on where you live? Evidence from Hurricane Katrina* (NBER Working Paper No. 24822). Cambridge, MA: National Bureau of Economic Research.

Doblhammer, G., Van den Berg, G. J., & Lumey, L. H. (2013). A re-analysis of the long-term effects on life expectancy of the Great Finnish Famine of 1866–68. *Population studies, 67*(3), 309-322.

Dowd, J. B., & Hamoudi, A. (2014). Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality. *International Journal of Epidemiology*, 43(4), 983-988.

Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979–1985. *Social Science & Medicine, 42*(1), 47-57.

Everett, B. G., Rehkopf, D. H., & Rogers, R. G. (2013). The nonlinear relationship between education and mortality: an examination of cohort, race/ethnic, and gender differences. *Population Research and Policy Review, 32*(6), 893-917.

Feldman, J. J., Makuc, D. M., Kleinman, J. C., & Cornoni-Huntley, J. (1989). National trends in educational differentials in mortality. *American Journal of Epidemiology, 129*(5), 919-933.

Finkelstein, A., Gentzkow, M., & Williams, H. L. (2019). *Place-based drivers of mortality: Evidence from migration* (NBER Working Paper No. 25975). Cambridge, MA: National Bureau of Economic Research.

Frisvold, D., & Golberstein, E. (2011). School quality and the education–health relationship: Evidence from Blacks in segregated schools. *Journal of Health Economics, 30*(6), 1232-1245.

Glied, S., & Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography, 45*(3), 741-761.

Goldin, C. (1998). America's graduation from high school: The evolution and spread of secondary schooling in the twentieth century. *The Journal of Economic History, 58*(2), 345-374.

Goldin, C., & Lleras-Muney, A. (2019). XX> XY?: The changing female advantage in life expectancy. *Journal of Health Economics, 67*, 102224.

Goldring, T., Lange, F., & Richards-Shubik, S. (2016). Testing for changes in the SES-mortality gradient when the distribution of education changes too. *Journal of Health Economics, 46*, 120-130.

Hayward, M. D., Hummer, R. A., & Sasson, I. (2015). Trends and group differences in the association between educational attainment and US adult mortality: Implications for understanding education's causal influence. *Social Science & Medicine, 127*, 8-18.

Hummer, R. A., & Hernandez, E. M. (2013). The effect of educational attainment on adult mortality in the United States. *Population Bulletin, 68*(1), 1.

Kitagawa, E. M., & Hauser, P. M. (1973). *Differential mortality in the United States: A study in socioeconomic epidemiology.* Cambridge, Harvard University Press.

Kraemer, S. (2000). The fragile male. *BMJ, 321*(7276), 1609-1612.

Kuznets, S., Miller, A. R., & Easterlin, R. A. (1960). Volume II (Analysis of Economic Change). In *Population Redistribution and Economic Growth. United States, 1870-1950*. Independence Square, Philadelphia: The American Philosophicla Society.

Lager, A. C. J., & Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences, 109*(22), 8461-8466.

Lauderdale, D. S. (2001). Education and survival: Birth cohort, period, and age effects. *Demography, 38*(4), 551-561.

Lindeboom, M., Portrait, F., & Van den Berg, G. J. (2010). Long-run effects on longevity of a nutritional shock early in life: the Dutch Potato famine of 1846–1847. *Journal of Health Economics, 29*(5), 617-629.

Link, B. G., Northridge, M. E., Phelan, J. C., & Ganz, M. L. (1998). Social epidemiology and the fundamental cause concept: on the structuring of effective cancer screens by socioeconomic status. *The Milbank Quarterly, 76*(3), 375-402.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, 80-94.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *The Review of Economic Studies, 72*(1), 189-221.

Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health: A hierarchical approach. *Demography, 40*(2), 309-331.

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly, 36*, 521-536.

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs, 27*(2), 350-360.

Montez, J. K., Hayward, M. D., & Zajacova, A. (2019). Educational disparities in adult health: US States as Institutional Actors on the Association. *Socius, 5*, 2378023119835345.

Montez, J. K., Hummer, R. A., & Hayward, M. D. (2012). Educational attainment and adult mortality in the United States: A systematic analysis of functional form. *Demography, 49*(1), 315-336.

Montez, J. K., Hummer, R. A., Hayward, M. D., Woo, H., & Rogers, R. G. (2011). Trends in the educational gradient of US adult mortality from 1986 through 2006 by race, gender, and age group. *Research on Aging, 33*(2), 145-171.

Montez, J. K., Zajacova, A., Hayward, M. D., Woolf, S. H., Chapman, D., & Beckfield, J. (2019). Educational disparities in adult mortality across US states: how do they differ, and have they changed since the mid-1980s? *Demography, 56*(2), 621-644.

Novosad, P., Rafkin, C., & Asher, S. (2020). Mortality change among less educated Americans.

Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., . . . Goldman, D. P. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs, 31*(8), 1803-1813.

Pappas, G., Queen, S., Hadden, W., & Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *New England Journal of Medicine, 329*(2), 103-109.

Preston, S. H., & Elo, I. T. (1995). Are educational differentials in adult mortality increasing in the United States? *Journal of Aging and Health, 7*(4), 476-496.

Preston, S. H., & Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography, 43*(4), 631-646.

Ross, C. E., Masters, R. K., & Hummer, R. A. (2012). Education and the gender gaps in health and mortality. *Demography, 49*(4), 1157-1183.

Rostron, B. L., Arias, E., & Boies, J. L. (2010). Education reporting and classification on death certificates in the United States. *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research*, (151), 1-21.

Sasson, I. (2016). Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010. *Demography, 53*(2), 269-293.

Stephens Jr, M., & Yang, D.-Y. (2014). Compulsory education and the benefits of schooling. *American Economic Review, 104*(6), 1777-1792.

Stinson, S. (1985). Sex differences in environmental sensitivity during growth and development. *American Journal of Physical Anthropology, 28*(S6), 123-147.

Van den Berg, G. J., Pinger, P. R., & Schoch, J. (2016). Instrumental variable estimation of the causal effect of hunger early in life on health later in life. *The Economic Journal*, *126*(591), 465-506.

Williams, D. R., & Jackson, P. B. (2005). Social sources of racial disparities in health. *Health Affairs*, *24*(2), 325-334.

Table 1. Summary statistics for whites born in the 48 states

	Full Sample	Male	Female
Panel A. Birth cohorts 1870-1915			
Longevity	76.20 (13.14)	73.98 (12.65)	79.10 (13.21)
Years of education	9.10 (3.17)	8.99 (3.28)	9.25 (3.01)
Year of birth	1896	1896	1897
Year of death	1972	1970	1976
Region (%)			
Northeast	12.24	12.57	11.80
Midwest	55.34	55.35	55.33
South	26.60	26.50	26.73
West	5.82	5.58	6.14
Observations	5,369,280	3,041,601	2,327,679
Panel B. Birth cohorts 1906-1915 surviving to age 35			
Longevity	75.85 (13.97)	72.79 (13.60)	79.63 (13.48)
Years of education	10.04 (2.96)	9.95 (3.05)	10.16 (2.83)
Year of birth	1910	1910	1910
Year of death	1986	1983	1990
Region (%)			
Northeast	11.99	12.26	11.67
Midwest	52.01	52.50	51.41
South	26.70	26.30	27.20
West	9.30	8.95	9.73
Observations	1,362,469	753,127	609,342
Panel C. State-level measures			
Relative Teachers' Wages	0.99	0.99	0.99
Length of Term	164.46	164.46	164.46
Pupil Teacher Ratio	30.74	30.74	30.74
Child Mortality per 1000 live births	185.29	185.29	185.29
Number of Physicians per 1000 population	1.21	1.21	1.21
Number of Nurses per 1000 population	0.52	0.52	0.52
Per Capita Income (Thousands in 1929 US dollars)	0.48	0.48	0.48
Observations	96	48	48

Note: Descriptive statistics were calculated based on whites born in the 48 states. In parentheses are standard deviations. We computed state specific child mortality (the number of deaths per 1000 live births) as the fraction of children that died among the number of children women ages 16-45 ever had based on the 1910 full census. We also calculated the average state-specific number of physicians/surgeons and nurses per 1000 population based on the occupation variable using the 1910 and 1920 full census. State-level per capita income is the average per capita income between 1900 and 1920.

Table 2. Regression results of longevity and education by gender.

	(1)	(2)	(3)	(4)	(5)
Panel A. Full Sample					
Education	0.44***	0.44***	0.42***	0.42***	0.44***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Female	6.75***	6.75***	6.75***	6.75***	6.61***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)
Observations	1,362,469	1,362,469	1,362,469	1,362,469	1,362,469
Adjust-R	0.0678	0.0678	0.0697	0.0697	0.066
Panel B. Male					
Education	0.44***	0.44***	0.40***	0.40***	0.44***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	753,127	753,127	753,127	753,127	753,127
Adjust-R	0.0097	0.0098	0.0125	0.0127	0.0115
Panel C. Female					
Education	0.43***	0.43***	0.43***	0.43***	0.44***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	609,342	609,342	609,342	609,342	609,342
Adjust-R	0.0083	0.0083	0.0100	0.0100	0.0089
State fixed effects	No	No	Yes	Yes	Yes
Cohort fixed effects	No	Yes	Yes	Yes	Yes
State-specific linear trends	No	No	No	Yes	Yes
Weights	No	No	No	No	Yes

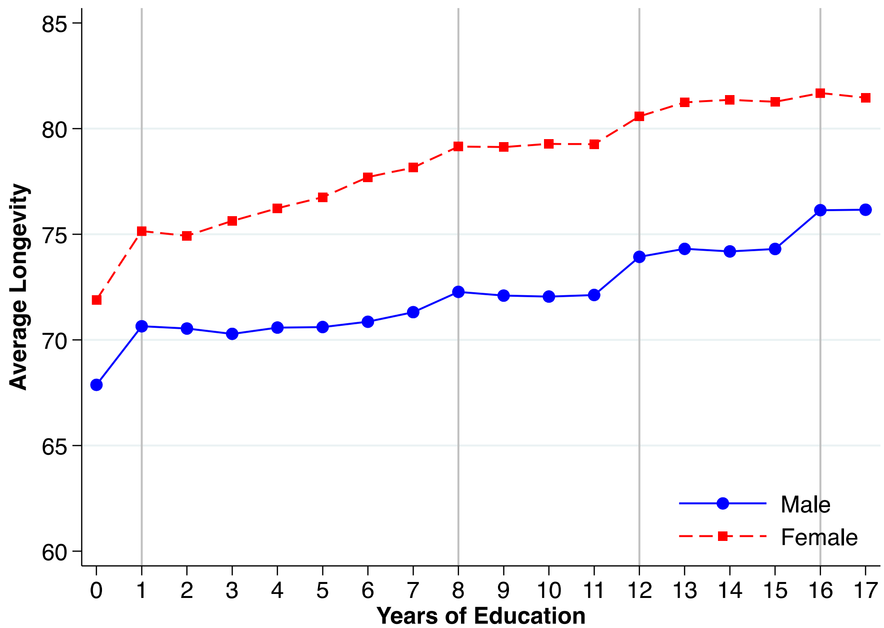
Note: Regression sample include whites born in the 48 states between 1906 and 1915 and conditional on being alive at age 35. N=1,362,469. Column 5 shows weighted estimates with state population as weights, making the sample representative of the US population in 1940. All estimates are from a linear regression model with Huber-White robust standard errors clustered at the state-of-birth level. *p < .05, **p < .01, ***p < .001.

Table 3. Estimates of the education gradient in alternative data sets, with and without truncation.

	CenSoc-Numident: Death year∈[1988, 2005]	Census-Tree: Death year∈[1988, 2005]	Census-Tree: Death year ∈[1941, 2005]	Census-Tree: Death year ∈[1941, 2019]
	(1)	(2)	(3)	(4)
Panel A. Full Sample				
Education	0.0882***	0.1042***	0.3500***	0.4176***
	(0.0035)	(0.0044)	(0.0188)	(0.0197)
Female	1.1618***	1.5379***	5.7276***	6.7515***
	(0.0199)	0.1042***	(0.1008)	(0.0930)
Observations	1,134,687	627,200	1,283,183	1,362,469
Adjust-R	0.1833	0.1905	0.0571	0.0697
Panel C. Male				
Education	0.0991***	0.1080***	0.3469***	0.4038***
	(0.0047)	(0.0055)	(0.0231)	(0.0247)
Observations	455,230	292,464	730,001	753,127
Adjust-R	0.1347	0.1649	0.0109	0.0127
Panel B. Female				
Education	0.0788***	0.0992***	0.3491***	0.4324***
	(0.0036)	(0.0047)	(0.0183)	(0.0193)
Observations	679,457	334,736	553,182	609,342
Adjust-R	0.1671	0.1638	0.0133	0.0100

Note: Analytic samples from CenSoc-Numident and Census-Tree data include whites born 1906-1915 in the 48 states. CenSoc-Numident data only include deaths from 1988 to 2005. All regressions include state-of-birth dummies, birth cohort dummies, and state-of-birth specific time trends. Standard errors are clustered at the state-of-birth level. *p < .05, **p < .01, ***p < .001.

Panel A. Education and mean longevity



Panel B. Distribution of longevity, by education levels

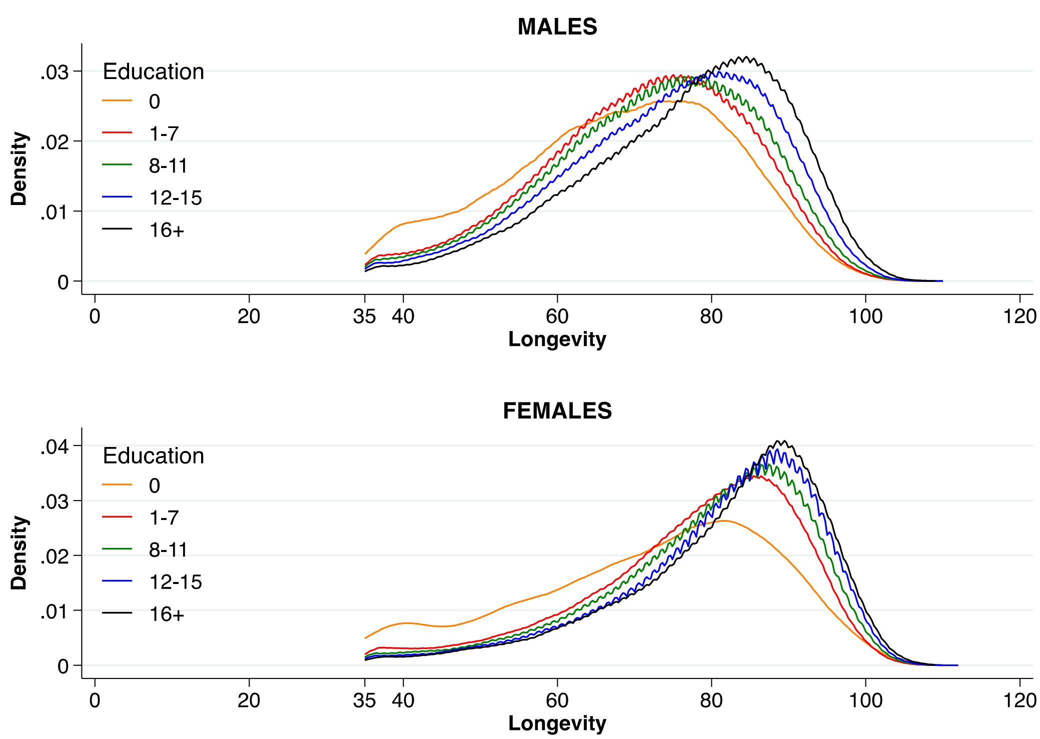
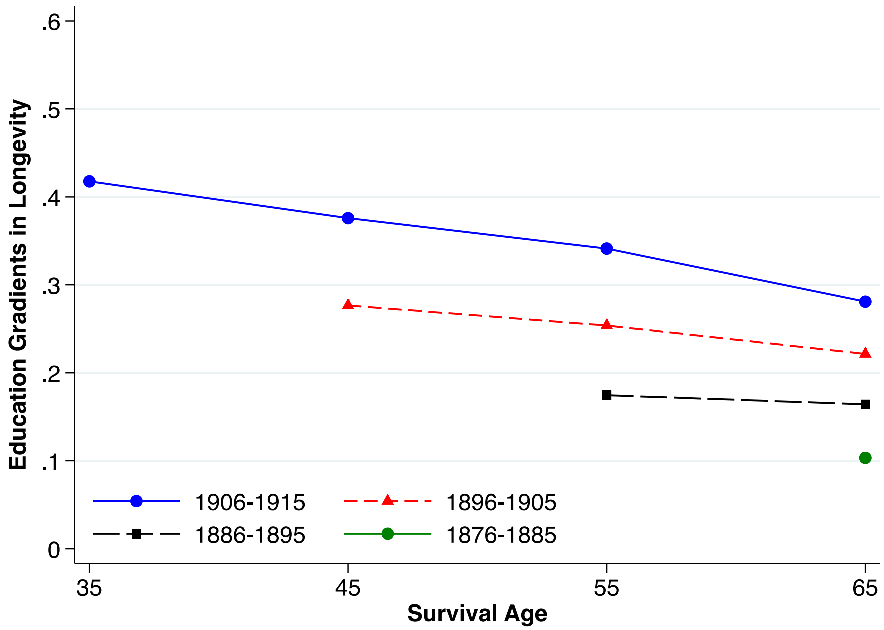


Fig. 1 Longevity for the 1906-1915 birth cohort by gender and by educational levels. Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. See text for details. N=1,362,469.

Panel A. Levels



Panel B. Logs

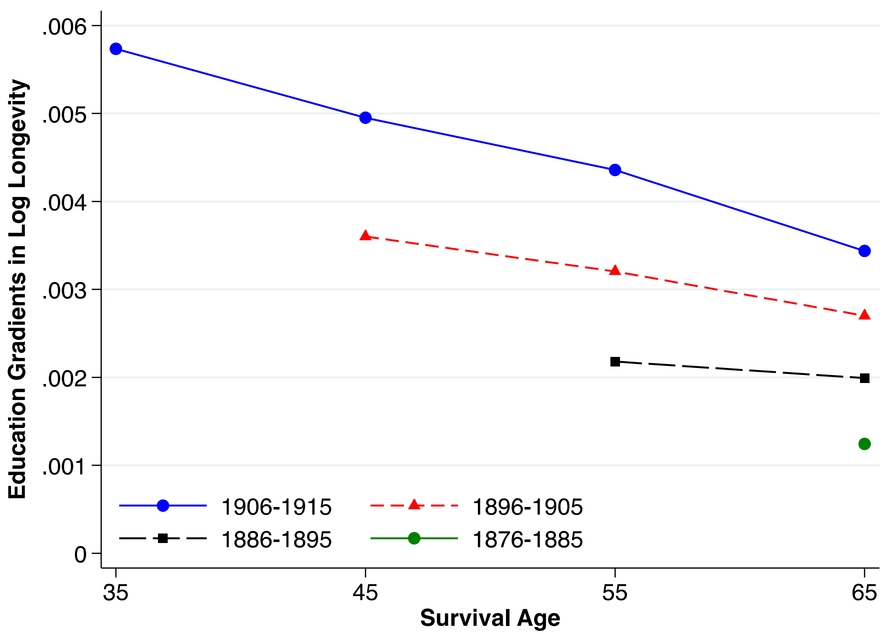
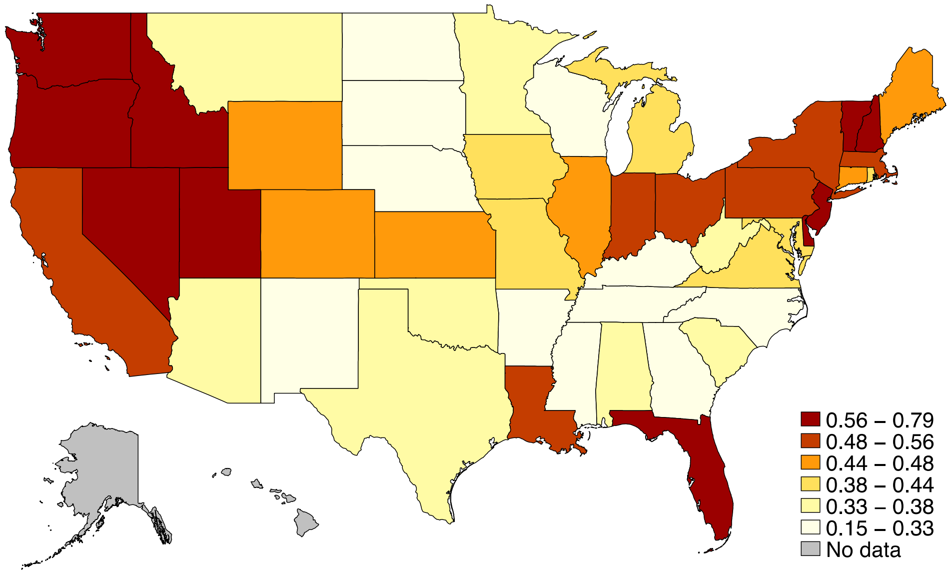


Fig. 2 Associations by age and birth cohort, across ten-year birth cohorts, conditional on being alive at age 35, 45, 55, and 65.Both figures show the coefficient of education, from regressions controlling for gender, birth cohort dummies, state of birth dummies, and state of birth specific time trends. Each estimate comes from a separate regression. We always restrict the sample so that all the individuals in a given cohort are restricted to have survived to the same age.

Panel A. For MALES



Panel B. FEMALES

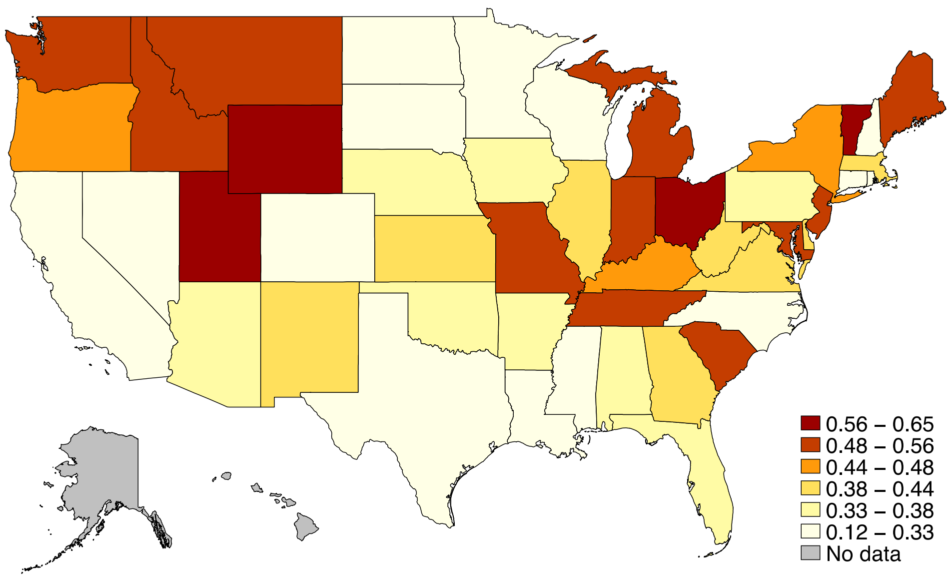


Fig. 3 Education gradients by state-of-birth, for the 1906-1915 cohort who were being alive at 35. The figure shows the association between education and longevity for each state. Specifically, we estimate a regression of longevity on years of education controlling for birth cohort and gender. The figure shows the coefficients for education for each state and gender. These coefficients are statistically significant for each state and they are different from each other. Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. See text for details. N=1,362,469.

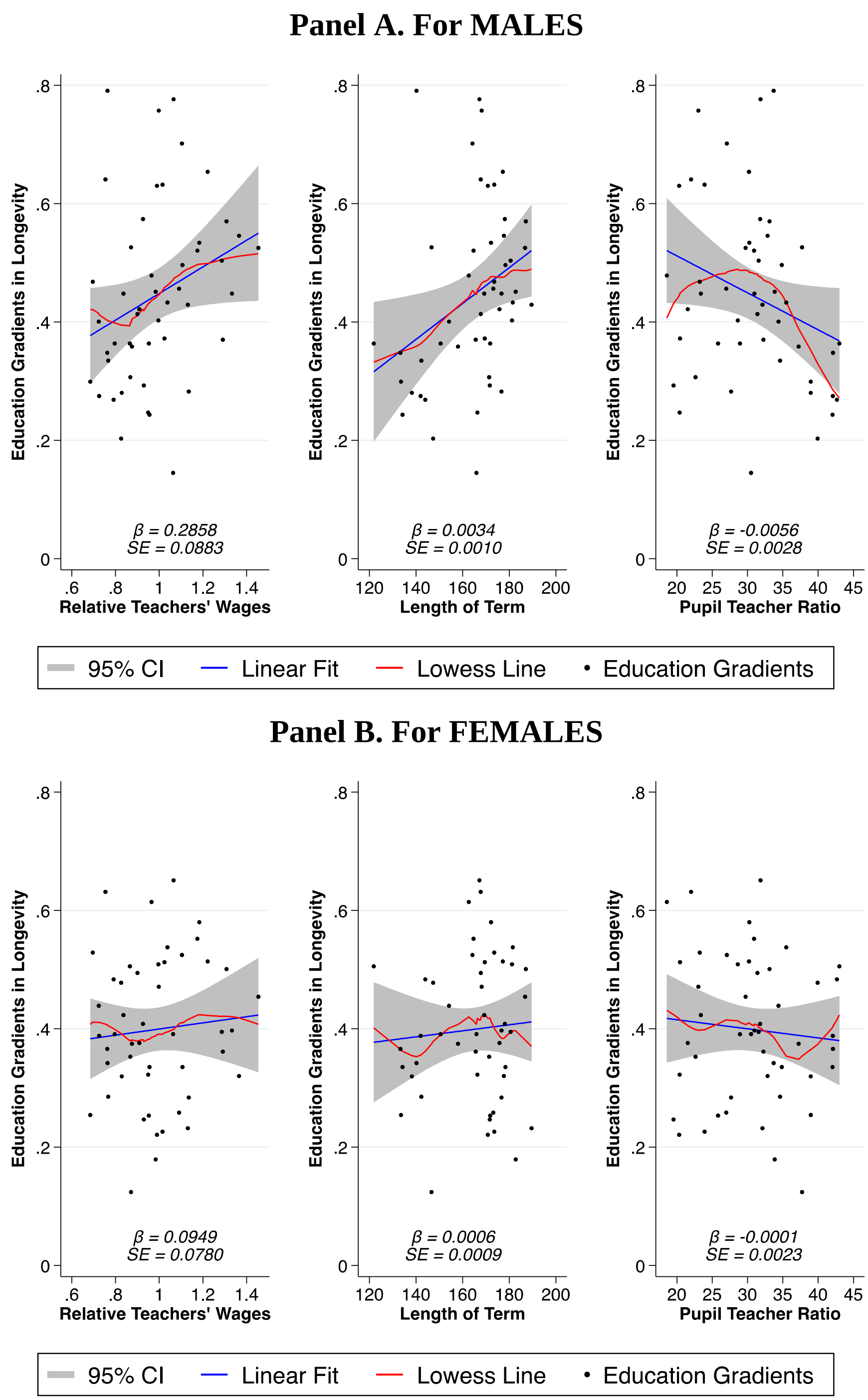
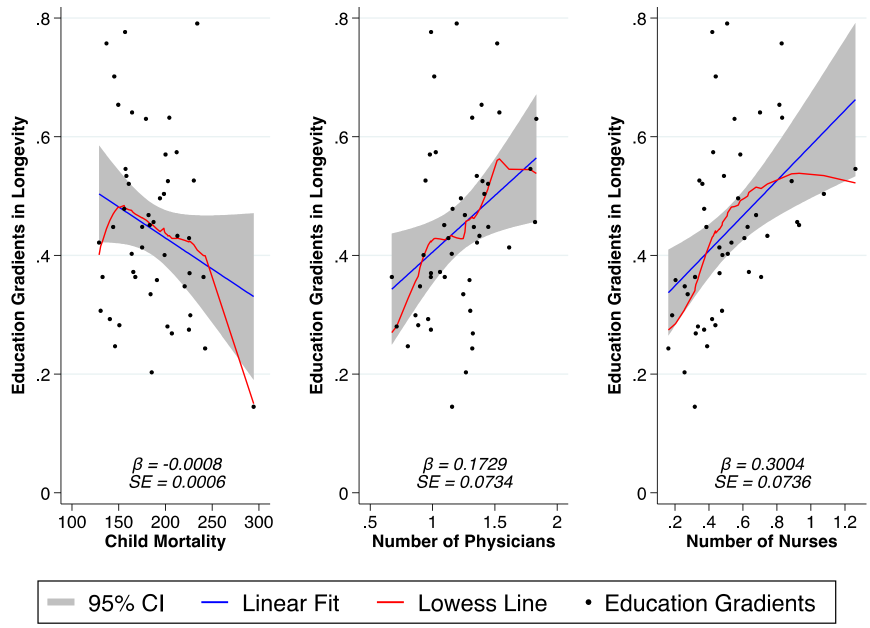


Fig. 4 Education gradients and state-level quality of schooling. Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. A point in the figure shows the state-level coefficient of education plotted against various measures of school quality. Coefficients (β) and standard errors (*SE*) are from a weighted OLS regression of education gradients on quality measures, with the weights equal to the inverse of standard errors of estimated gradients.

Panel A. For MALES



Panel B. For FEMALES

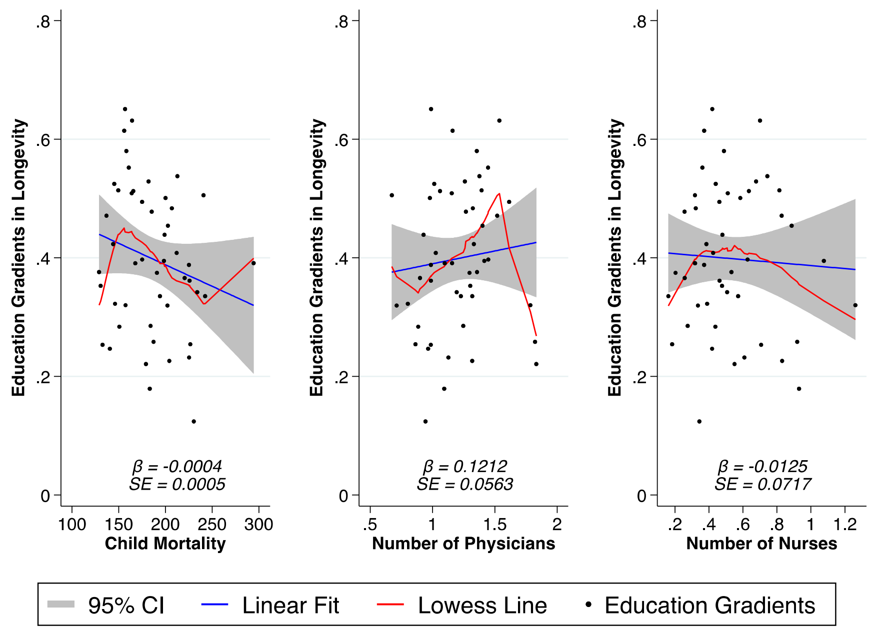


Fig. 5 Education gradients and state-level proxies of health resources. Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. A point in the figure shows the state-level coefficient of education. Coefficients (β) and standard errors (SE) are from a weighted OLS regression of education gradients on heath measures, with the weights equal to the inverse of standard errors of estimated gradients.

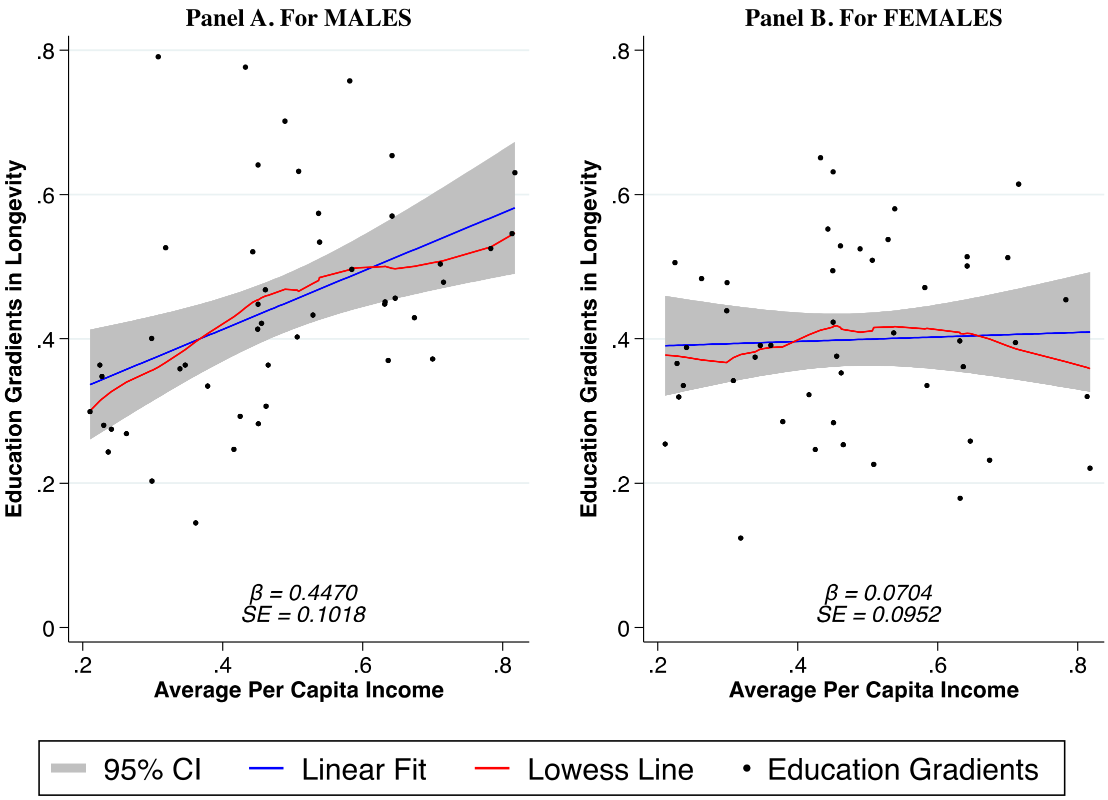


Fig. 6 Education gradients and per capita income. Sample from the Census-Tree data for estimating longevity returns includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. State-level per capita income is the average per capita income between 1900 and 1920. A point in the figure shows the state-level coefficient of education plotted against per capita income. Coefficients (β) and standard errors (*SE*) are from a weighted OLS regression of education gradients on state-level per capita income, with the weights equal to the inverse of standard errors of estimated gradients.

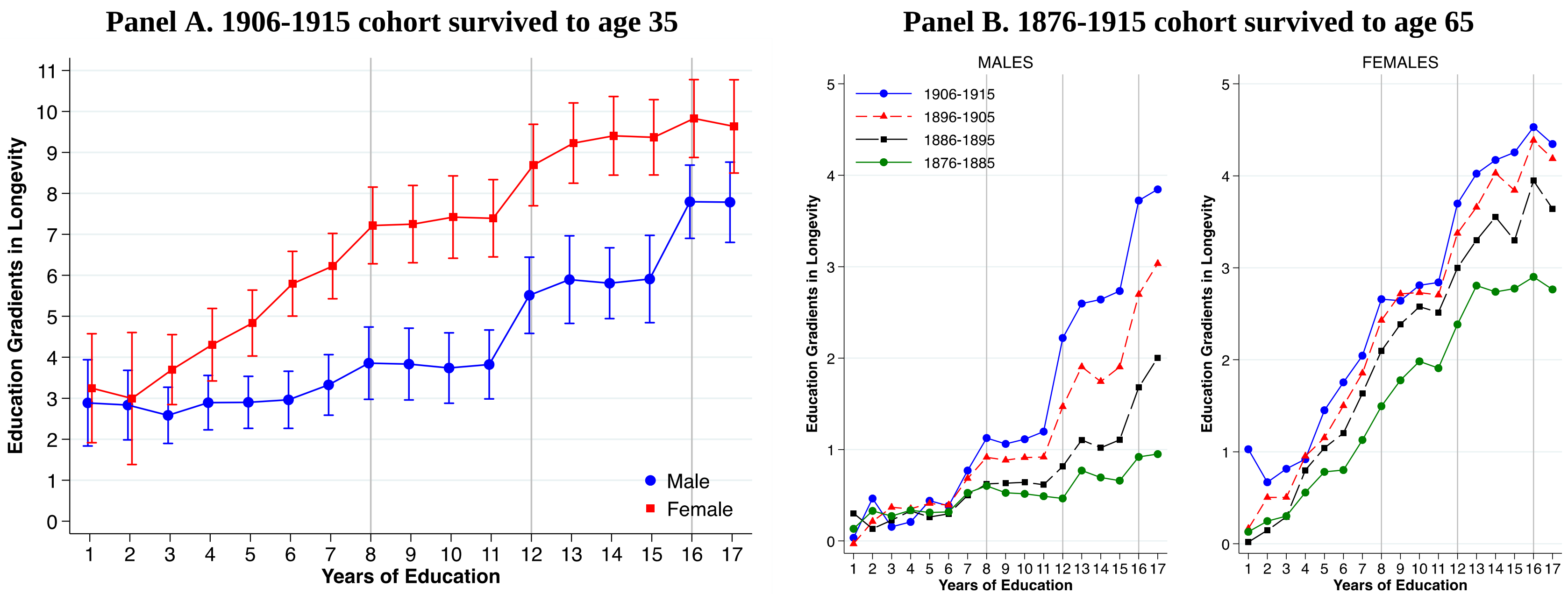


Fig. 7 Non-parametric estimates of the education-longevity relationship by gender and birth cohort. Panel A reports the coefficients from a regression of age at death on dummies for each single year of school, controlling for state of birth dummies, year of birth dummies and state-of-birth specific linear trend. The excluded category is 0 years of school. The sample includes only whites born 1906-1915 in the 48 states of the US who survived to age 35. Panel B reports the coefficients from a regression of age at death on dummies for each single year of school, controlling for state of birth dummies, year of birth dummies and state-of-birth specific linear trend. The excluded category is 0 years of school. The 1906-1915/1986-1905/1886-1895/1876-1885 samples includes only whites born in the 48 states of the US who survived to age 65.

Appendix Tables and Figures

Not for publication

Table A1. Descriptive statistics of 1906-1915 analytic sample in comparison to the 1940 full-count census

	Census-Tree linked data						1940 full-count data					
	ALL		Male		Female		ALL		Male		Female	
Year of birth	N	Edu	%	Edu	%	Edu	N	Edu	%	Edu	%	Edu
1906	151,429	9.80	54.17	9.70	45.83	9.92	1,710,279	9.56	49.87	9.44	50.13	9.68
1907	150,258	9.85	54.50	9.74	45.50	9.98	1,691,639	9.68	49.70	9.56	50.30	9.79
1908	145,850	9.91	54.70	9.81	45.30	10.03	1,888,340	9.74	49.28	9.63	50.72	9.85
1909	140,974	9.96	54.89	9.85	45.11	10.08	1,687,964	9.91	49.72	9.81	50.28	10.01
1910	137,724	9.99	55.35	9.89	44.65	10.12	1,993,185	9.86	48.81	9.72	51.19	9.99
1911	132,801	10.05	55.68	9.95	44.32	10.17	1,852,255	9.94	49.66	9.83	50.34	10.05
1912	134,239	10.11	55.74	10.02	44.26	10.22	1,934,501	9.98	49.10	9.87	50.90	10.08
1913	131,043	10.17	56.15	10.08	43.85	10.30	1,907,268	10.08	49.53	9.97	50.47	10.19
1914	130,421	10.27	56.70	10.19	43.30	10.37	1,955,931	10.16	49.00	10.06	51.00	10.25
1915	128,264	10.36	56.47	10.30	43.53	10.44	2,007,208	10.24	48.72	10.15	51.28	10.31

Note: To be consistent with 1940 full census, estimates, estimates from the Census-Tree data includes whites born 1906-1915 in the 48 states, not restricting to those being alive at 35. N=1,383,003.

Table A2. Representativeness the sample data by state-of-birth: comparing the distribution of observations in the 1940 census and in the census-tree data for the 1906-1915 cohort.

		Census-Tree Data			1940 Full Census		
FIPS	State	All (%)	Male (%)	Female (%)	All (%)	Male (%)	Female (%)
1	Alabama	0.36	0.35	0.37	1.58	1.57	1.59
4	Arizona	0.33	0.32	0.35	0.37	0.39	0.36
5	Arkansas	2.48	2.41	2.57	1.2	1.19	1.21
6	California	0.22	0.22	0.22	5.97	6.12	5.83
8	Colorado	0.39	0.36	0.42	0.91	0.92	0.91
9	Connecticut	0.77	0.79	0.74	1.45	1.44	1.46
10	Delaware	0.02	0.02	0.02	0.2	0.2	0.2
12	Florida	0.08	0.07	0.09	1.21	1.19	1.22
13	Georgia	2.65	2.6	2.72	1.8	1.8	1.79
16	Idaho	1.52	1.43	1.64	0.43	0.46	0.41
17	Illinois	8.2	8.29	8.08	6.59	6.52	6.66
18	Indiana	5.91	5.95	5.85	2.68	2.71	2.66
19	Iowa	5.8	5.88	5.71	1.94	1.96	1.93
20	Kansas	0.88	0.86	0.91	1.36	1.35	1.36
21	Kentucky	4.84	4.76	4.93	2.05	2.07	2.02
22	Louisiana	0.1	0.09	0.1	1.36	1.36	1.35
23	Maine	1.36	1.44	1.27	0.63	0.64	0.63
24	Maryland	1.05	1.11	0.99	1.34	1.37	1.32
25	Massachusetts	2.35	2.47	2.21	3.47	3.37	3.57
26	Michigan	4.78	4.85	4.7	4.25	4.31	4.19
27	Minnesota	0.77	0.72	0.83	2.24	2.26	2.21
28	Mississippi	0.18	0.17	0.19	0.94	0.94	0.95
29	Missouri	5.97	5.91	6.03	2.9	2.86	2.93
30	Montana	0.91	0.89	0.92	0.45	0.48	0.42
31	Nebraska	3.01	3	3.02	1.01	1.01	1.01
32	Nevada	0.12	0.12	0.12	0.1	0.11	0.09
33	New Hampshire	0.14	0.13	0.15	0.38	0.38	0.38
34	New Jersey	1.28	1.31	1.25	3.48	3.45	3.51
35	New Mexico	0.29	0.28	0.3	0.41	0.41	0.41
36	New York	3.7	3.74	3.64	11.52	11.23	11.8
37	North Carolina	4.78	4.73	4.83	2.21	2.2	2.22
38	North Dakota	1.6	1.61	1.6	0.5	0.52	0.48
39	Ohio	8.76	8.98	8.5	5.45	5.43	5.47
40	Oklahoma	0.66	0.64	0.68	1.75	1.74	1.76
41	Oregon	0.29	0.28	0.3	0.91	0.94	0.88
42	Pennsylvania	1.29	1.26	1.32	7.94	7.94	7.94
44	Rhode Island	0.36	0.37	0.36	0.59	0.56	0.62
45	South Carolina	1.54	1.47	1.62	0.93	0.93	0.93
46	South Dakota	1.55	1.53	1.59	0.47	0.47	0.46
47	Tennessee	3.73	3.68	3.79	2.03	2.01	2.05
48	Texas	0.53	0.52	0.54	4.9	4.91	4.9
49	Utah	2.87	2.66	3.13	0.43	0.43	0.43
50	Vermont	0.72	0.73	0.7	0.27	0.28	0.26
51	Virginia	0.55	0.52	0.58	1.76	1.79	1.72
53	Washington	1.92	1.98	1.84	1.45	1.5	1.4
54	West Virginia	3.23	3.24	3.22	1.45	1.47	1.43
55	Wisconsin	4.73	4.85	4.57	2.51	2.55	2.47
56	Wyoming	0.43	0.41	0.46	0.22	0.23	0.2
	Northeast	11.97	12.24	11.64	29.73	29.29	30.17
	Midwest	51.96	52.43	51.39	31.9	31.95	31.83
	South	26.78	26.38	27.24	26.71	26.74	26.66
	West	9.29	8.95	9.70	11.65	11.99	11.34
	N	1,383,003	766,030	616,973	19,098,225	9,442,561	9,655,664

Note: Sample includes only whites born 1906-1915 in the 48 states of the US.

Table A3. Summary statistics for whites born 1906-1915 in the 48 states and died 1988-2005 using Berkeley's CenSoc-Numident data and our Census-Tree data.

	CenSoc-Numident		Census-Tree	
	Male	Female	Male	Female
Longevity	82.71 (5.23)	84.75 (5.32)	84.23 (5.16)	86.12 (5.29)
Min Longevity	72	72	73	73
Max Longevity	99	99	99	99
Years of education	10.79 (3.12)	10.54 (2.89)	10.26 (3.06)	10.26 (2.80)
Year of birth (%)	1912	1912	1911	1910
1906	30.79	5.12	7.82	10.41
1907	37.46	5.75	8.58	10.62
1908	23.24	6.06	9.09	10.62
1909	8.51	6.65	9.53	10.48
1910	1.82	7.29	9.88	10.18
1911	2.16	8.22	10.17	9.86
1912	2.56	11.40	10.91	10.02
1913	3.12	12.95	11.08	9.53
1914	3.59	15.76	11.48	9.31
1915	5.42	20.79	11.47	8.96
Region (%)				
Northeast	30.79	29.42	12.42	11.44
Midwest	37.46	35.55	53.73	51.43
South	23.24	28.13	24.01	27.09
West	8.51	6.9	9.84	10.03
Observations	455,230	679,457	292,464	334,736

Note: In parentheses are standard deviations.

The CenSoc-Numident data include 6,824,036 individuals, among which 50% are females, and 93.19% are whites. Since this data only provides death records between 1988 and 2005, the mean age at death is 75.75, with a standard deviation equals to 9.07 and a range (47, 121).

We first restricted the CenSoc-Numident dataset to those born 1906-1915 in 48 U.S. states (N= 1,214,847) and then merged it to the 1940 Full Census. All 1,214,847 observations from the CenSoc-Numident can be matched to the census. However, there are disagreements with respect to race between Census and CenSoc data; 94.75% (N= 1,151,073) identify themselves as white based on Census, while 86.63% (N=1,052,530) people consider themselves as white in both Census and CenSoc. To be consistent with the definition of our Census-Tree data, we restricted to whites using the self-reported race from the Census only. We further excluded those missing years of education and included 1,134,687 in the analyses.

Table A4. **OLS of log of age at death on a continuous measure of education for the 1906-1915 birth cohort who were alive at age 35.**

	(1)	(2)	(3)	(4)
Panel A. Full Sample				
Education	0.0060***	0.0060***	0.0057***	0.0057***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Female	0.0912***	0.0912***	0.0913***	0.0912***
	(0.0013)	(0.0013)	(0.0014)	(0.0014)
State fixed effects	No	No	yes	yes
Cohort fixed effects	No	Yes	yes	yes
State-specific linear trends	No	No	no	yes
Observations	1,362,469	1,362,469	1,362,469	1,362,469
Adjust-R	0.0580	0.0581	0.0597	0.0598
AIC	-539,299	-539,317	-541,783	-541,909
BIC	-539,263	-539,172	-541,650	-541,751
Panel B. Male				
Education	0.0061***	0.0061***	0.0056***	0.0056***
	(0.0003)	(0.0003)	(0.0004)	(0.0004)
State fixed effects	No	No	yes	yes
Cohort fixed effects	No	Yes	yes	yes
State-specific linear trends	No	No	no	yes
Observations	753,127	753,127	753,127	753,127
Adjust-R	0.0084	0.0084	0.0110	0.0111
AIC	-260,846	-260,874	-262,863	-263,011
BIC	-260,823	-260,747	-262,748	-262,896
Panel C. Female				
Education	0.0058***	0.0059***	0.0058***	0.0058***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
State fixed effects	no	No	yes	Yes
Cohort fixed effects	no	yes	yes	Yes
State-specific linear trends	no	No	no	Yes
Observations	609,342	609,342	609,342	609,342
Adjust-R	0.0074	0.0074	0.0089	0.0089
AIC	-280,644	-280,657	-281,589	-281,653
BIC	-280,621	-280,532	-281,476	-281,472

Note: Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. Standard errors are clustered at the state-of-birth level. **p* < .05, ***p* < .01, ****p* < .001.

Table A5: Education gradients in longevity by state of birth. Top and bottom ten states.

Top 10 States			Bottom 10 States		
State	Education Gradients	10p-90p increases	State	Education Gradients	10p-90p increase
Panel A. Levels					
1.Utah	0.7240	4.3441	39. Minnesota	0.3201	2.2404
2. Oregon	0.6487	4.5410	40. Kentucky	0.3193	2.5542
3. Vermont	0.6369	3.1847	41. Texas	0.3132	2.5059
4. Idaho	0.6296	3.7776	42. North Carolina	0.2977	2.3816
5. Florida	0.6220	5.5983	43. Mississippi	0.2928	2.6354
6. Washington	0.6023	4.2163	44. Wisconsin	0.2834	1.7004
7. Ohio	0.5519	3.3117	45. Arkansas	0.2815	1.9703
8. New Jersey	0.5433	4.3462	46. North Dakota	0.2796	1.6777
9. Delaware	0.5340	4.8062	47. South Dakota	0.2729	1.6372
10. Indiana	0.5337	2.6685	48. New Mexico	0.2566	2.3093
Panel B. Logs					
1. Utah	0.0100	0.0600	39. Georgia	0.0044	0.0348
2. Vermont	0.0088	0.0441	40. Minnesota	0.0043	0.0303
3. Oregon	0.0088	0.0616	41. Texas	0.0042	0.0334
4. Florida	0.0086	0.0777	42. North Carolina	0.0041	0.0325
5. Idaho	0.0086	0.0515	43. Wisconsin	0.0039	0.0234
6. Washington	0.0084	0.0585	44. North Dakota	0.0038	0.0230
7. Delaware	0.0079	0.0714	45. Arkansas	0.0038	0.0267
8. Ohio	0.0077	0.0462	46. Mississippi	0.0037	0.0335
9. Wyoming	0.0075	0.0447	47. New Mexico	0.0037	0.0331
11. New Jersey	0.0073	0.0586	48. South Dakota	0.0036	0.0215

Notes: Analytic sample includes whites born between 1806 and 1915 in the 48 states who were alive at age 35 (N=1,362,469). We estimate a linear regression model stratified by state of birth to estimate the state specific returns to education, and to report the increase in longevity when education increases from the 10th percentile to 90th percentile. Estimates of state-specific returns to education on longevity reported in the table adjust for gender, and birth cohort fixed effects.

Table A6. Returns to education on longevity and quality of schooling (similar to the Table 2 in Card and Kruger 1992 paper)

	Male				Female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Quality of Schooling								
Relative Teachers' Wages	0.2858**			0.2057	0.0949			0.1013
	(0.0883)			(0.1226)	(0.0780)			(0.1165)
Length of Term		0.0034**		0.0012		0.0006		0.0000
		(0.0010)		(0.0018)		(0.0009)		(0.0018)
Pupil Teacher Ratio			-0.0056*	-0.0029			-0.0001	0.0007
			(0.0028)	(0.0035)			(0.0023)	(0.0034)
Adjusted R^2	0.1677	0.1726	0.0611	0.1880	0.0101	-0.0121	-0.0217	-0.0328
P value of joint significance				0.0067				0.6830
Panel B: Public Health								
Child mortality	-0.0008			-0.0005	-0.0004			-0.0002
	(0.0006)			(0.0005)	(0.0005)			(0.0005)
Number of Physicians		0.1729*		0.0495		0.1212*		0.1447*
		(0.0734)		(0.0760)		(0.0563)		(0.0628)
Number of Nurses			0.3004***	0.2667**			-0.0125	-0.0890
			(0.0736)	(0.0832)			(0.0717)	(0.0757)
Adjusted R^2	0.0196	0.0882	0.2496	0.2453	-0.0097	0.0718	-0.0211	0.0604
P value of joint significance				0.0015				0.1268
Panel C: Economy								
Per Capita Income	0.4470***				0.0704			
	(0.1018)				(0.0952)			
Adjusted R^2	0.2800				-0.0097			

Note. Estimates are from a weighted OLS regression of education gradients on state-level measures, with the weights equal to the inverse of standard errors of estimated gradients. We computed state specific child mortality (the number of deaths per 1000 live births) as the fraction of children that died among the number of children women ages 16-45 ever had based on the 1910 full census. We also calculated the average state-specific number of physicians/surgeons and nurses per 1000 population based on the occupation variable using the 1910 and 1920 full census. State-level per capita income is the average per capita income between 1900 and 1920. N=48. *$p < .05$, **$p < .01$, ***$p < .001$.

Table A7. Education gradients in longevity and state-level measures

	Males	Females
Pupil Teacher Ratio	0.0001	0.0019
	(0.0039)	(0.0045)
Length of Term	0.0013	0.0017
	(0.0017)	(0.0021)
Relative Teachers' Wages	0.0721	0.1379
	(0.1264)	(0.1415)
Average Longevity	-0.0680**	-0.0863**
	(0.0227)	(0.0276)
Average Years of Education	0.1534***	0.0573
	(0.0263)	(0.0340)
Child Mortality per 1000 live births	-0.0002	-0.0006
	(0.0006)	(0.0007)
Number of Physicians per 1000 population	0.0845	0.1314*
	(0.0616)	(0.0637)
Number of Nurses per 1000 population	0.0940	-0.1470
	(0.1018)	(0.1267)
Per Capita Income (thousands in 1929 US dollars)	-0.2866	-0.2848
	(0.2578)	(0.3398)
Constant	3.5698*	6.3828**
	(1.5254)	(2.0974)
Observations	48	48
Adjusted R^2	0.5677	0.1844

Note: Estimates are from a weighted OLS regression of education gradient on state-level measures, with the weights equal to the inverse of standard errors of estimated gradients. We computed state specific child mortality (the number of deaths per 1000 live births) as the fraction of children that died among the number of children women ages 16-45 ever had based on the 1910 full census. We also calculated the average state-specific number of physicians/surgeons and nurses per 1000 population based on the occupation variable using the 1910 and 1920 full census. State-level per capita income is the average per capita income between 1900 and 1920. *$p < .05$, **$p < .01$, ***$p < .001$.

Table A8. Estimates from a non-parametric model.

	Males		Females	
	beta	95% CI	beta	95% CI
Years of Schooling (0 = reference)				
1	2.89	(1.83,3.94)	3.25	(1.92,4.58)
2	2.83	(1.99,3.68)	2.99	(1.38,4.60)
3	2.58	(1.90,3.27)	3.70	(2.84,4.55)
4	2.89	(2.23,3.56)	4.31	(3.42,5.19)
5	2.90	(2.27,3.54)	4.83	(4.03,5.64)
6	2.96	(2.27,3.66)	5.79	(5.00,6.58)
7	3.33	(2.59,4.06)	6.22	(5.43,7.02)
8	3.86	(2.97,4.74)	7.22	(6.28,8.15)
9	3.83	(2.96,4.71)	7.25	(6.31,8.19)
10	3.74	(2.88,4.60)	7.42	(6.42,8.43)
11	3.82	(2.98,4.66)	7.39	(6.45,8.33)
12	5.51	(4.58,6.44)	8.69	(7.70,9.68)
13	5.89	(4.82,6.96)	9.23	(8.25,10.21)
14	5.81	(4.94,6.67)	9.40	(8.44,10.36)
15	5.91	(4.85,6.97)	9.37	(8.45,10.29)
16	7.79	(6.90,8.69)	9.83	(8.88,10.78)
17+	7.78	(6.80,8.76)	9.63	(8.50,10.77)
State of birth	Yes		Yes	
Year of birth	Yes		Yes	
State-of-birth specific linear trend	Yes		Yes	
Constant	227.96	(220.27,235.64)	161.43	(154.88,167.99)
Observations	753,127		609,342	
Adjusted R^2	0.0139		0.0109	
AIC	6,058,507		4,892,274	
BIC	6,058,819		4,892,591	

Note: CI represents the Confidence Interval. The estimates are from a regression of age at death on dummies for each single year of school, controlling for state of birth dummies, year of birth dummies and state-of-birth specific linear trend. The excluded category is 0 years of school. The sample includes only whites born 1906-1915 in the 48 states of the US who survived to age 35. Standard errors are clustered by state-of-birth.

Table A9. Test for linearity for the 1906-1915 cohort

	Male		Female	
	F-stat	P-value	F-stat	P-value
sch2 - sch1 = sch3 - sch2	0.144	0.706	0.354	0.555
sch3 - sch2 = sch4 - sch3	1.544	0.220	0.014	0.905
sch4 - sch3 = sch5 - sch4	0.836	0.365	0.057	0.813
sch5 - sch4 = sch6 - sch5	0.077	0.782	2.007	0.163
sch6 - sch5 = sch7 - sch6	3.477	0.068	6.183	0.017
sch7 - sch6 = sch8 - sch7	0.785	0.380	6.322	0.015
sch8 - sch7 = sch9 - sch8	9.456	0.004	14.023	0.000
sch9 - sch8 = sch10 - sch9	0.252	0.618	1.038	0.313
sch10 - sch9 = sch11 - sch10	2.481	0.122	2.484	0.122
sch11 - sch10 = sch12 - sch11	56.595	0.000	43.088	0.000
sch12 - sch11 = sch13 - sch12	42.937	0.000	15.304	0.000
sch13 - sch12 = sch14 - sch13	3.001	0.090	2.849	0.098
sch14 - sch13 = sch15 - sch14	0.556	0.460	0.738	0.395
sch15 - sch14 = sch16 - sch15	42.928	0.000	2.914	0.094
sch16 - sch15 = sch17 - sch16	50.596	0.000	3.663	0.062
P-value of the joint test (linearity)	43.30	0.000	13.73	0.000

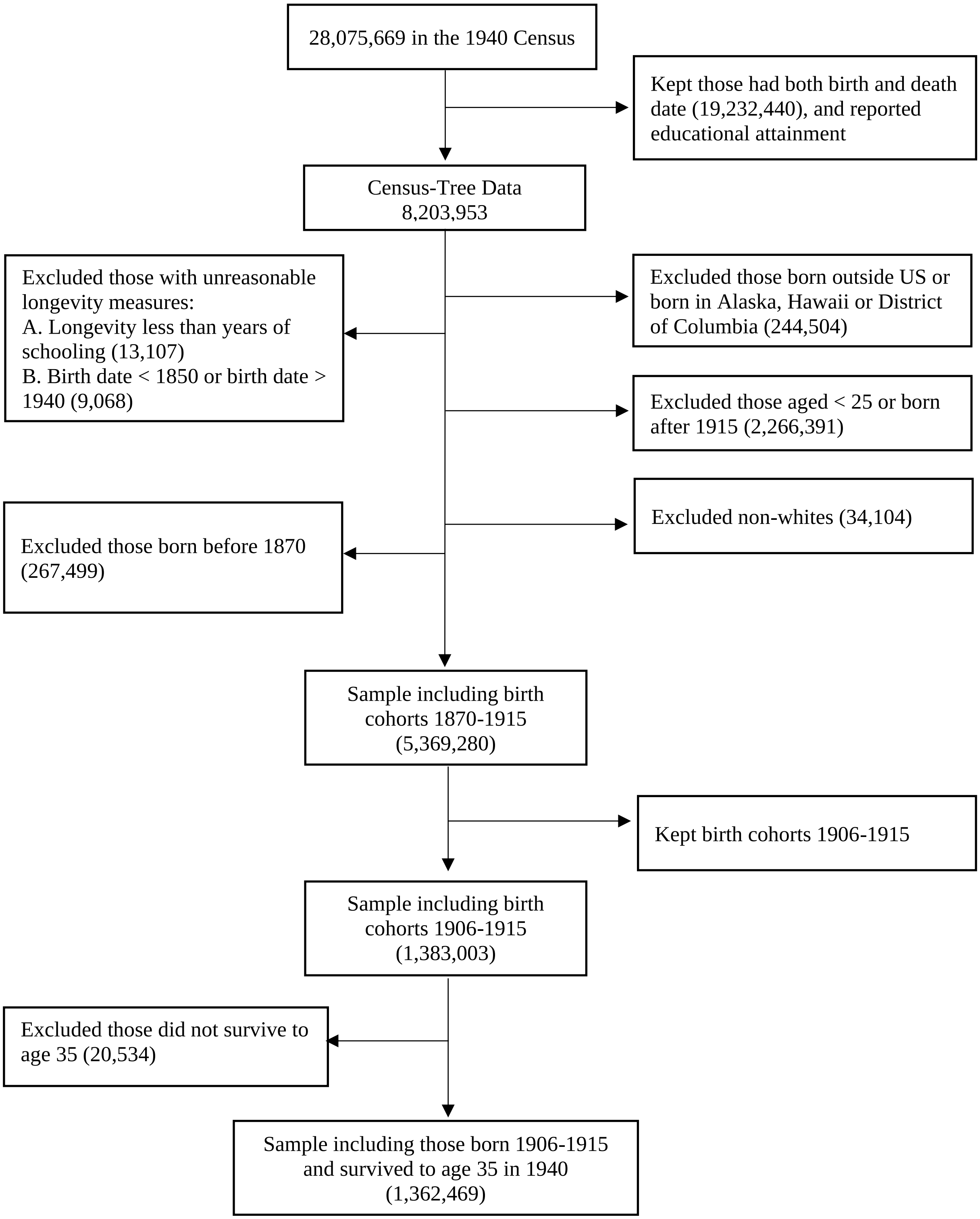
Note: The estimates are F test for coefficients from a regression of age at death on dummies for each single year of school, controlling for state of birth dummies, year of birth dummies and state-of-birth specific linear trend. The excluded category is 0 years of school. The sample includes only whites born 1906-1915 in the 48 states of the US who survived to age 35. Standard errors are clustered by state-of-birth.

Table A10. Fit measures for different models. OLS, various splines and fully non-parametric model

Models	Non-parametric	Linear	Spline
Panel A. Full Sample			
Adjusted R^2	0.0132	0.0121	0.0132
MSE	192.5309	192.7357	192.5371
AIC	11,033,276	11,034,744	11,033,320
BIC	11,033,700	11,035,083	11,033,647
Panel B. Male			
Adjusted R^2	0.0139	0.0127	0.0139
MSE	182.4615	182.6866	182.4652
AIC	6,058,507	6,059,408	6,058,506
BIC	6,058,819	6,059,546	6,058,713
Panel C. Female			
Adjusted R^2	0.0109	0.010	0.0108
MSE	179.6184	179.7709	179.6238
AIC	4,892,274	4,892,765	4,892,276
BIC	4,892,591	4,892,924	4,892,491

Note: Analytic sample includes whites born between 1906 and 1915 in the 48 states surviving to age 35. We included dummies of exact years of education (no education as the reference) in the non-parametric model, included a continuous measure of years of education in the linear model, and used the following knots for the spline regression model: years of education=1, 7, 8, 11, 12, 15, and 16. All models control for gender (not in the stratified analyses by gender), state of birth fixed effects, year of birth fixed effects, and state of birth specific linear trends. Standard errors are clustered at the state-of-birth level. MSE stands for Mean Squared Error, which is calculated from 10-fold cross validation. AIC denotes Akaike Information Criterion, and BIC represents Bayesian Information Criterion. Based on BIC measures, the spline model is the preferred model to depict the relationship between educational attainment and longevity.

Figure A1: Sample Flowchart



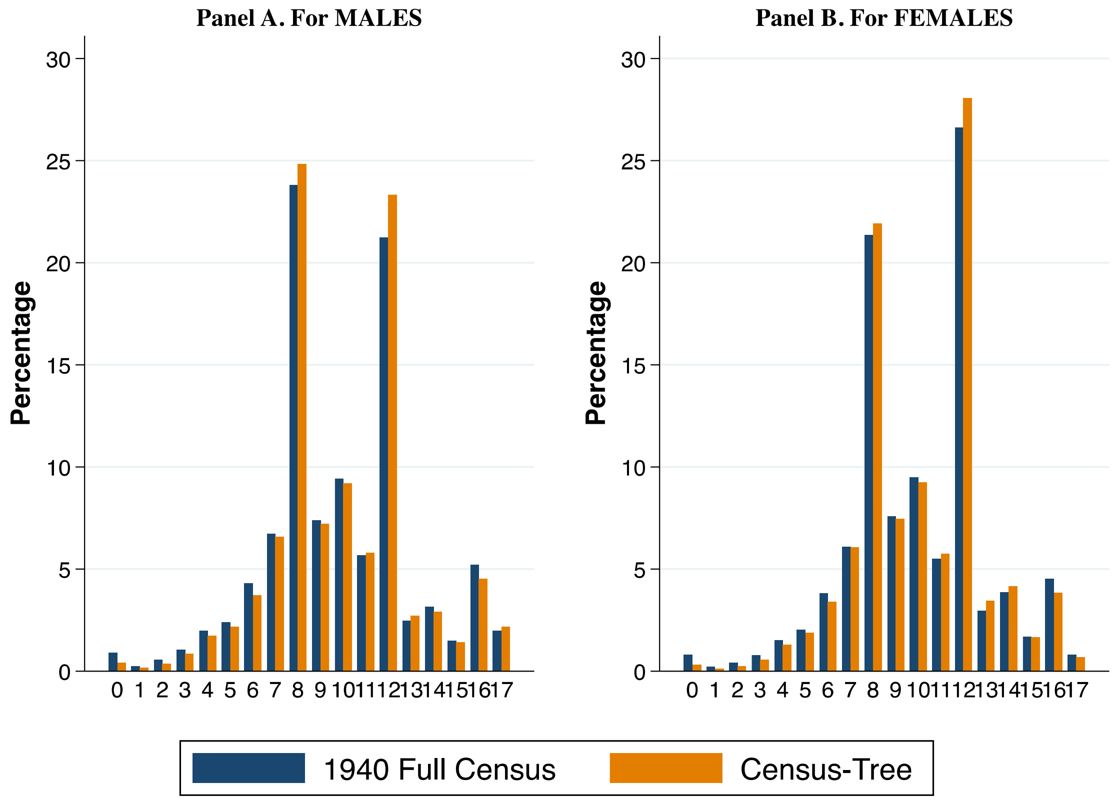


Fig. A2 Distribution of years of education in the 1906-1915 birth cohort. Matched 1940-family tree data. Note: To be consistent with 1940 full census, estimates, estimates from the Census-Tree data includes whites born 1906-1915 in the 48 states, not restricting to those being alive at 35. N=1,383,003.

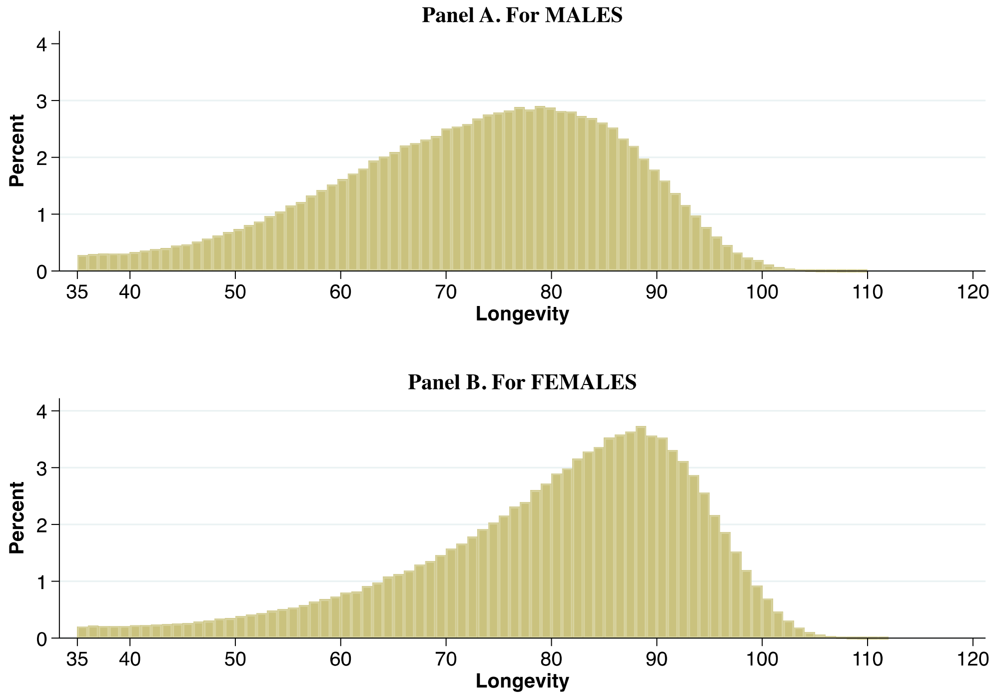


Fig. A3 Distribution of longevity for the 1906-1915 birth cohort by gender. The sample includes only whites born 1906-1915 in the 48 states of the US who survived to age 35.

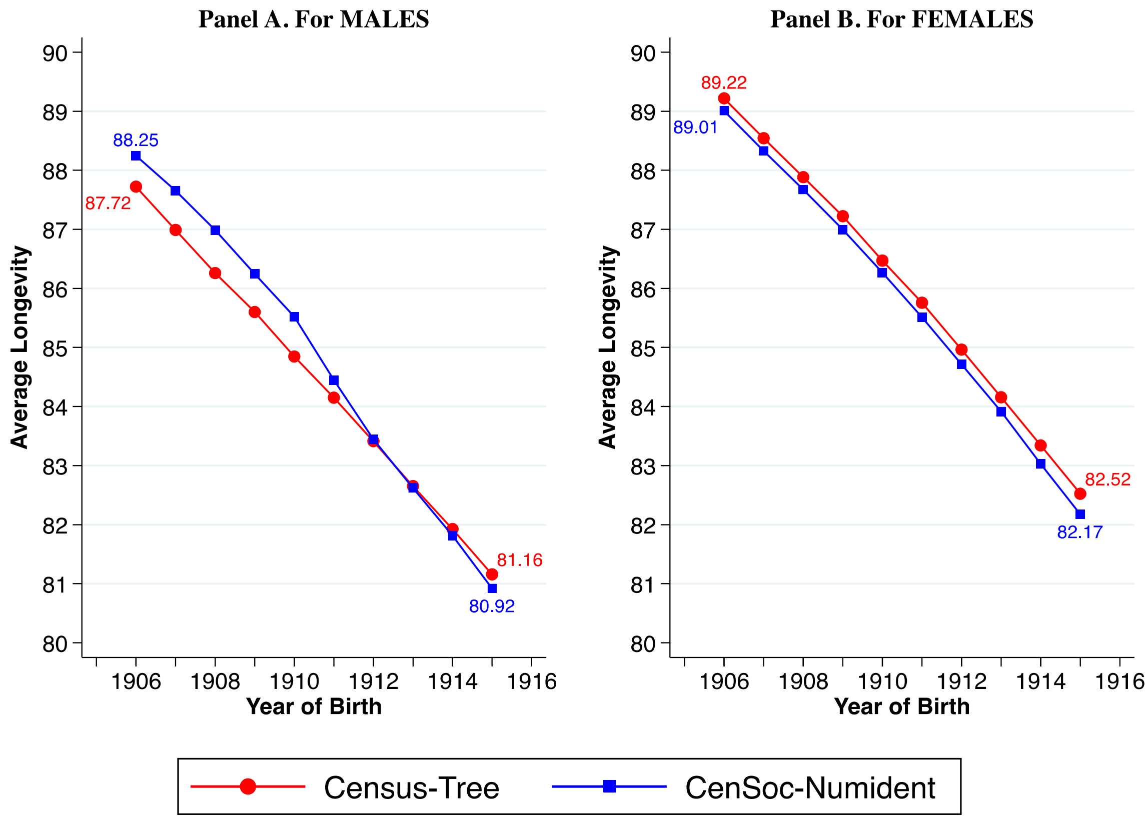


Fig. A4 Restrict to whites in CenSoc-Numident based on self-reported race in the 1940 Census. Estimates are based on CenSoc-Numident from UC Berkeley and Census-Tree data used in this study. The analytic sample include whites born 1906-1915 in the 48 states, and who died between 1988 and 2005. In the Censoc-Numident data and in our data the race is defined using the race reported in the 1940 census. There are 1,134,687 observations in CenSoc-Numident, and 627,200 observations in Census Tree. This figure shows a downwards trend in average longevity by birth cohorts. The primary reason is the survival bias. Those born in 1870 were 70 years old in the 1940 census, while those born in 1910 were only 30.

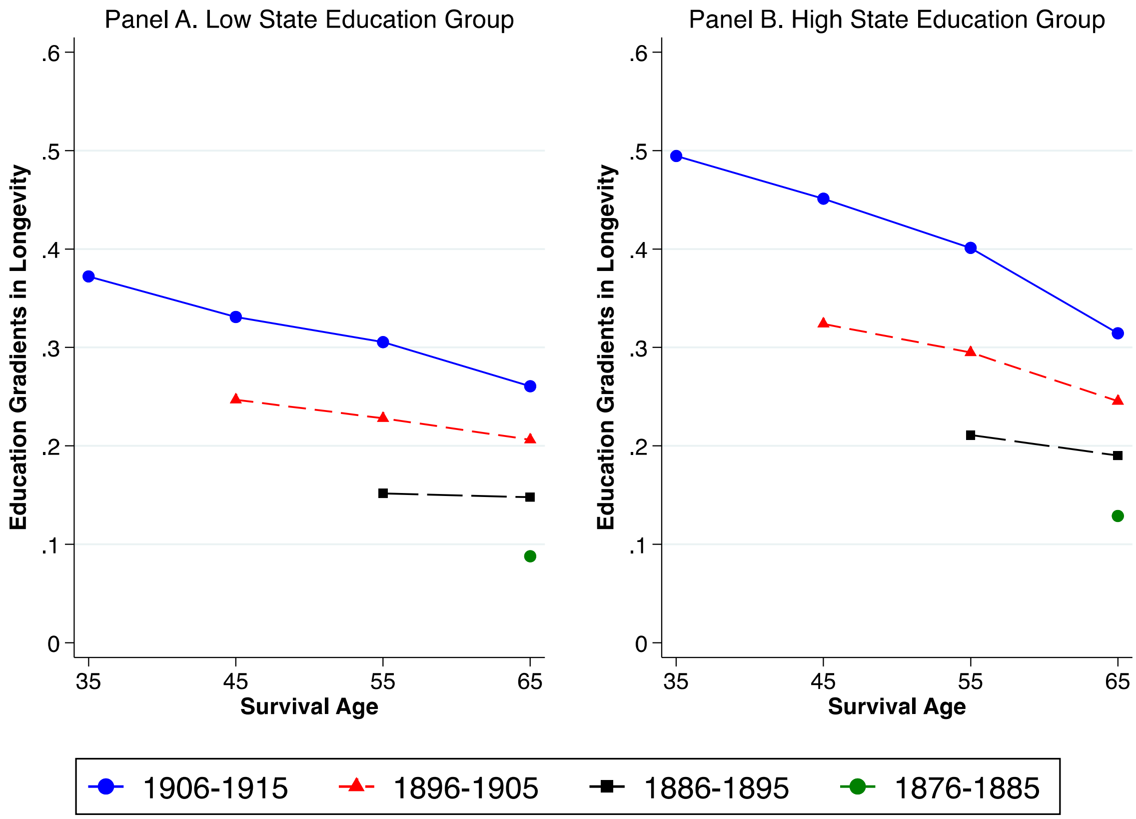


Fig. A5 Associations by age and birth cohort, across ten-year birth cohorts, conditional on being alive at age 35, 45, 55, and 65, and by quartiles of state-level average years of education. Both figures show the coefficient of education, from regressions controlling for gender, birth cohort dummies, state of birth dummies, and state of birth specific time trends. Each estimate comes from a separate regression. We always restrict the sample so that all the individuals in a given cohort are restricted to have survived to the same age. States with above-median average education are included in the high state education group.

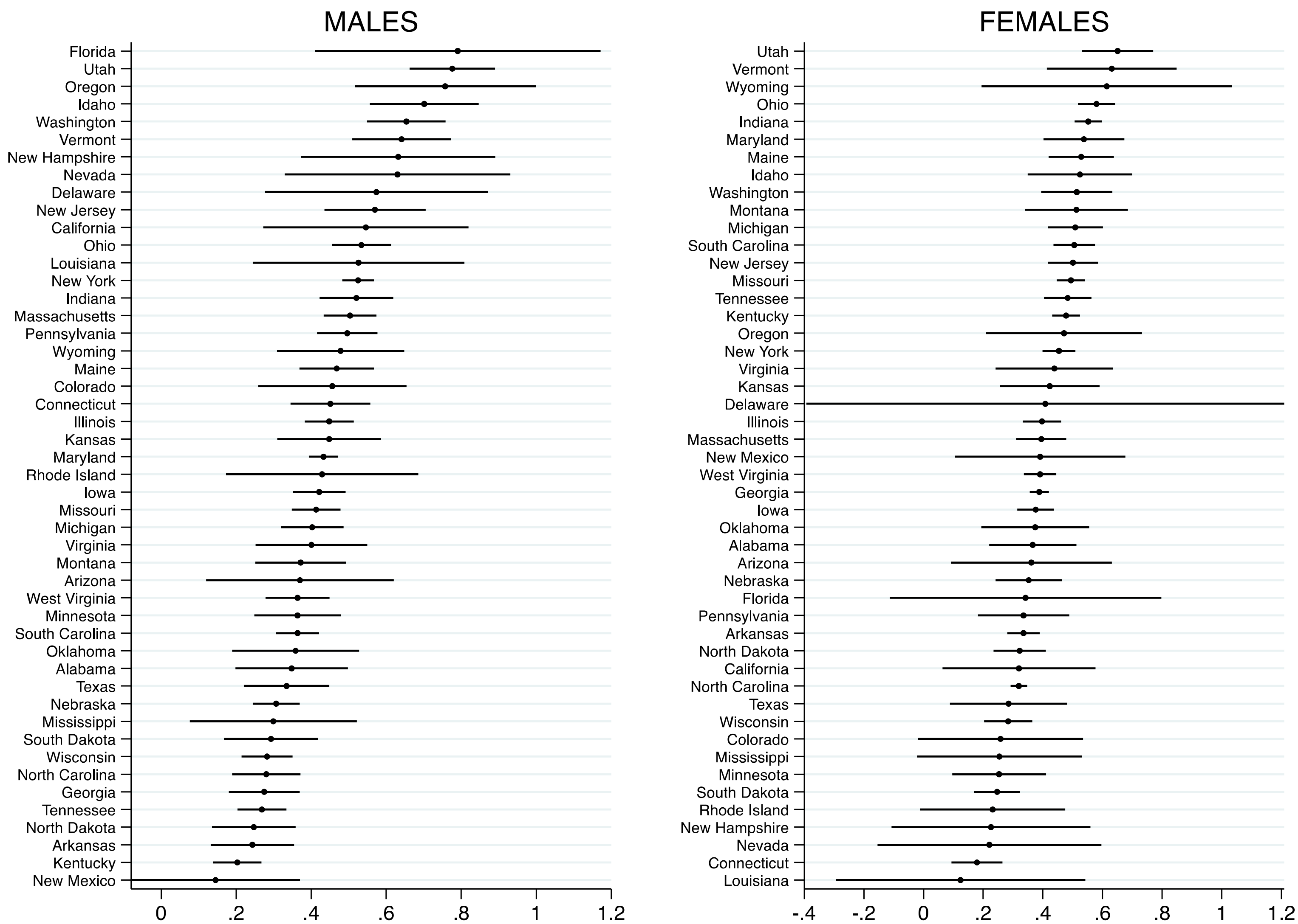


Fig. A6 Education gradient in longevity by gender and state of birth. The sample includes only whites born 1906-1915 in the 48 states of the US who survived to age 35.

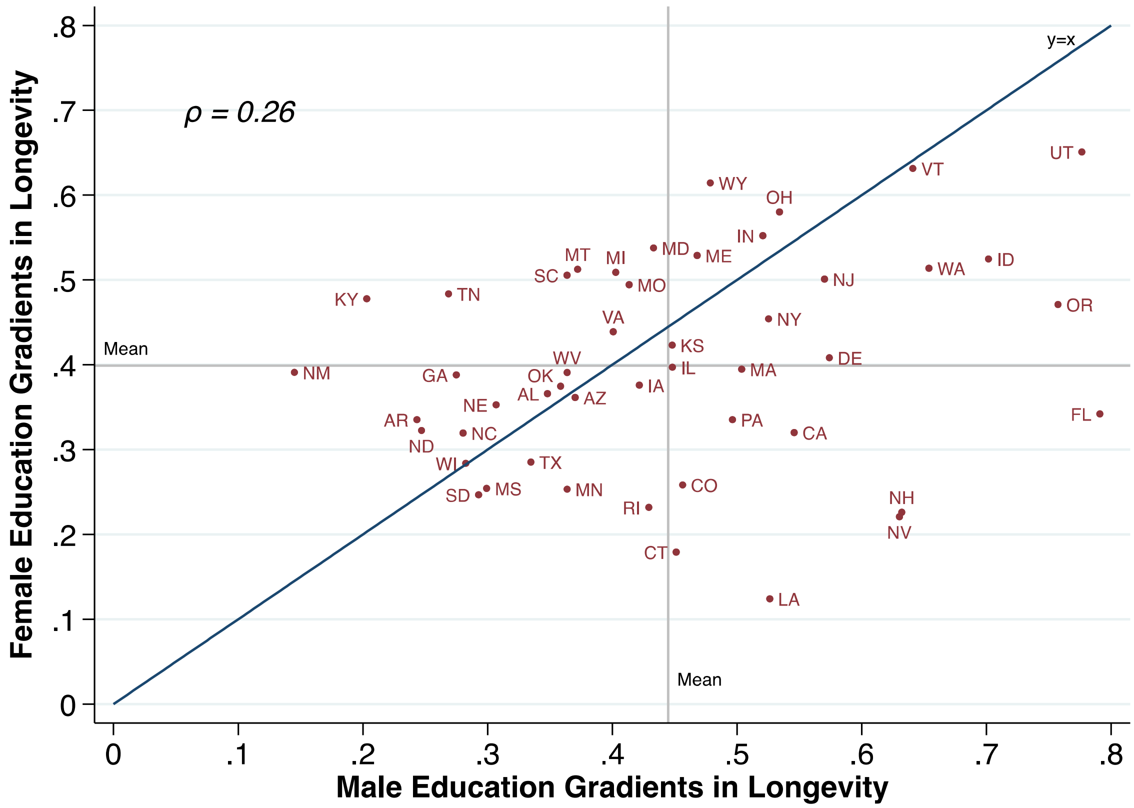


Fig. A7 Correlation of female and male returns to education by state for the 1906-1915 cohort. Matched 1940-family tree data. Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. See text for details. N=1,362,469. Coefficients for males and for females estimated separately for each state. Regressions control for year of birth. The correlation (ρ) between male and female coefficients is 0.26.

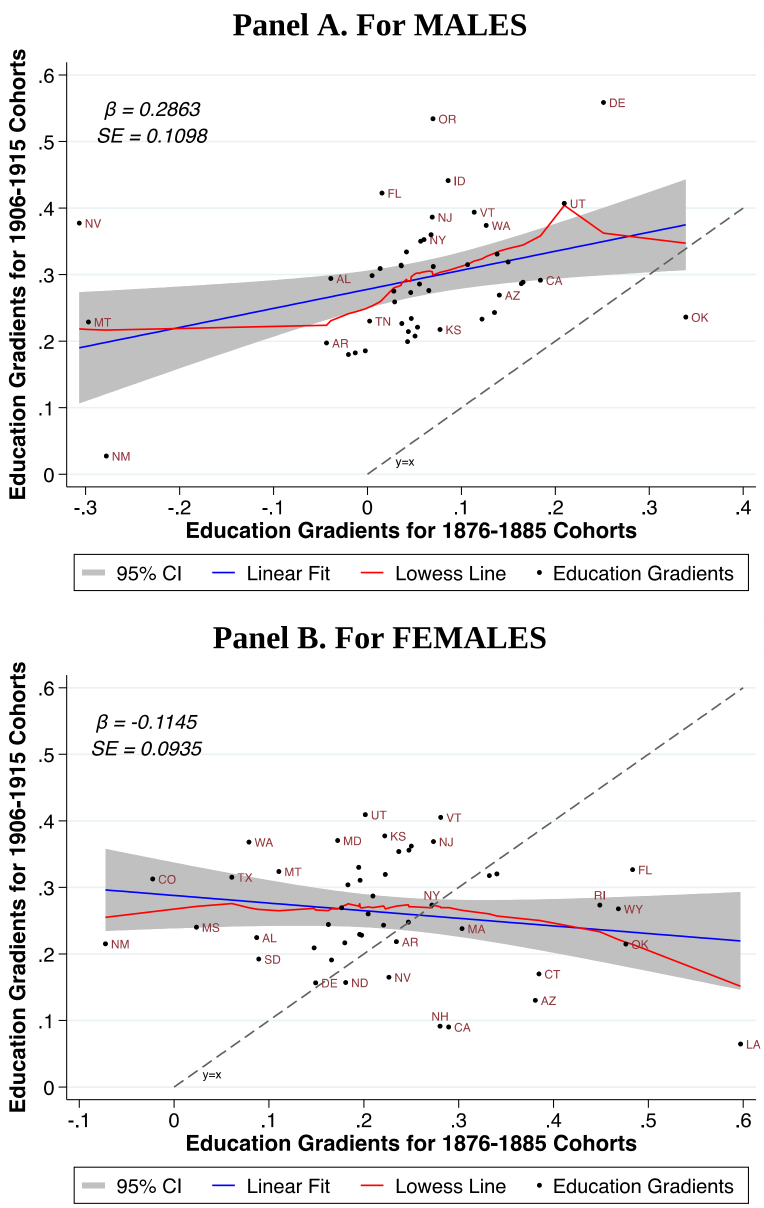
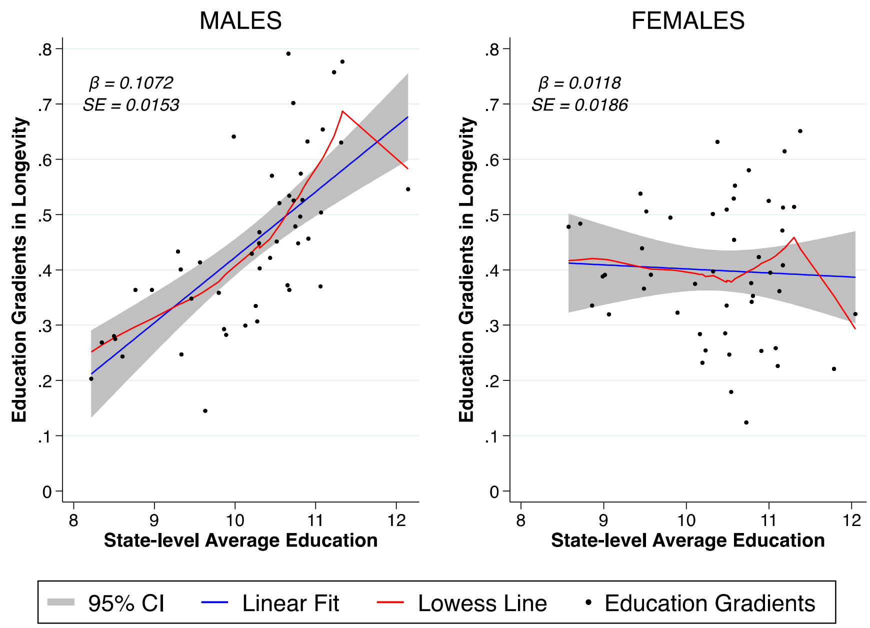


Fig. A8 Education gradients at age 65 by birth cohorts. Education gradients in longevity are estimated those born in 1906-1915 and 1876-1885. For both birth cohorts, analytic sample includes whites born in 48 states and survived to age 65. Coefficients (β) and standard errors (*SE*) are from a weighted OLS regression of education gradients for 1906-1915 cohorts on education gradients for 1876-1885 cohorts

Panel A. By education levels



Panel B. By longevity levels

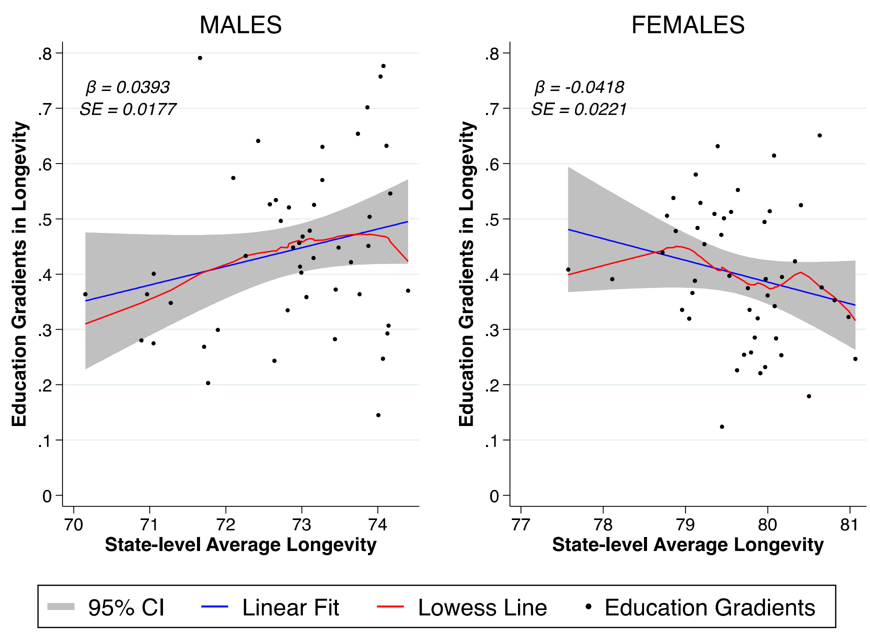
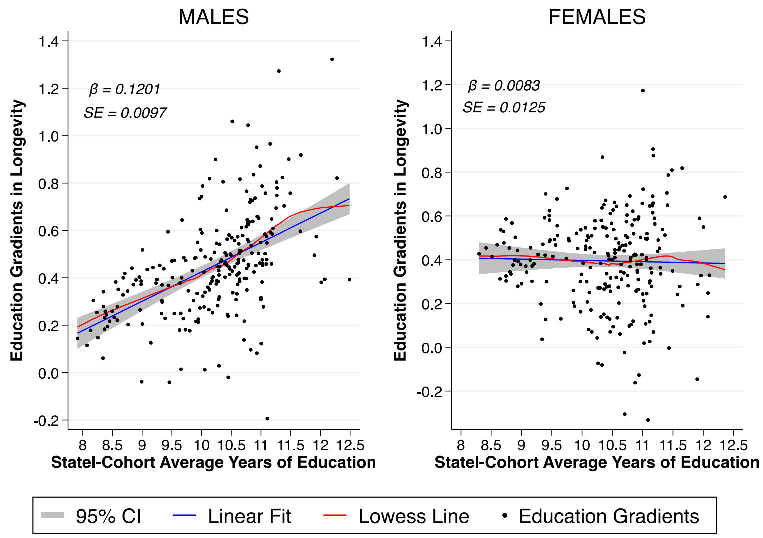


Fig. A9 Education gradients by level of education and longevity. Matched 1940-family tree data. Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. See text for details. N=1,362,469. Coefficients (β) and standard errors (*SE*) are from a weighted OLS regression of education gradients on state-level average education or longevity, with the weights equal to the inverse of standard errors of estimated gradients.

.

Panel A. By education levels



Panel B. By longevity levels

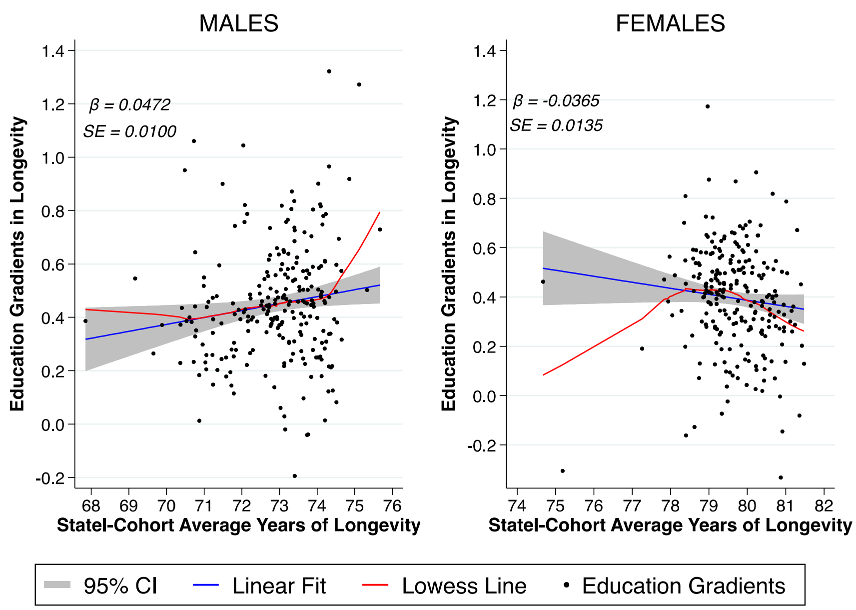


Fig. A10. Returns to education on longevity by level of education and longevity
Estimates by state of birth and cohort. Each point denotes returns to education for specific state*cohort cell (48 states by 5 birth cohort groups with two birth cohorts per group). Sample includes only whites born 1906-1915 in the 48 states of the US and survived to age 35. Coefficients (β) and standard errors (*SE*) are from a weighted OLS regression of education gradients on state-cohort level average education or longevity, with the weights equal to the inverse of standard errors of estimated gradients.

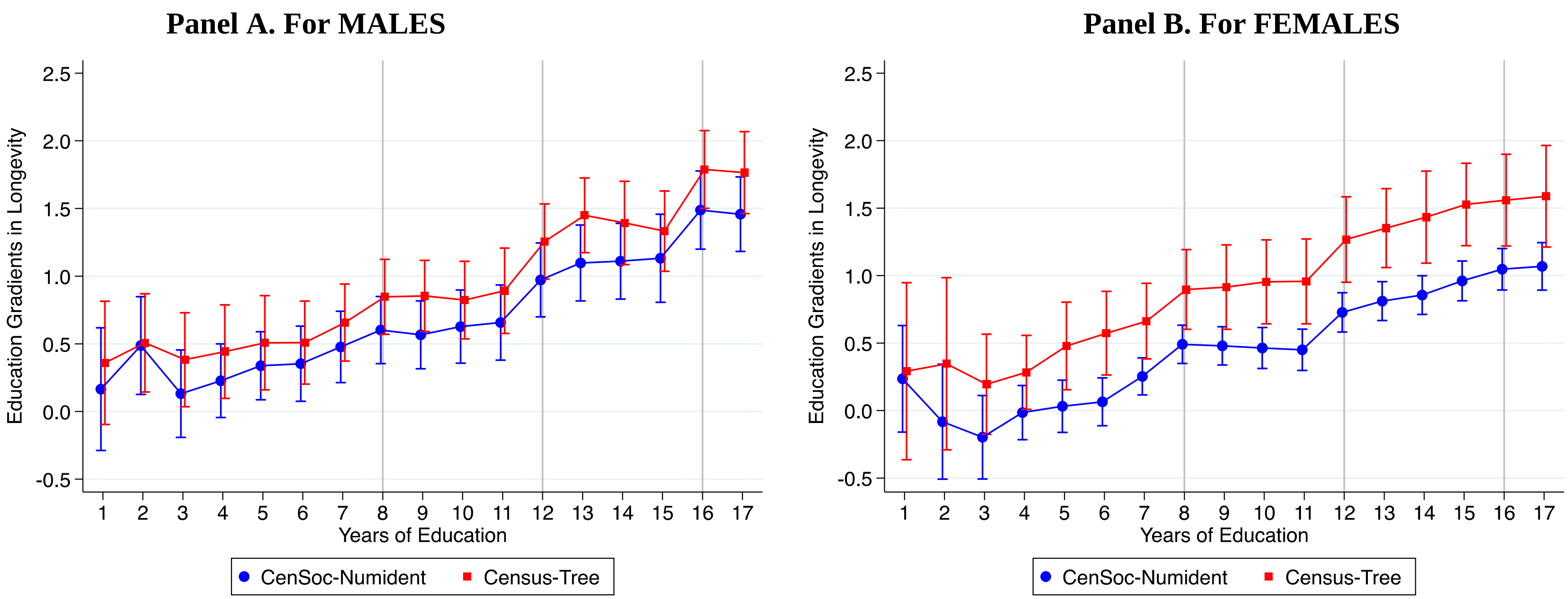
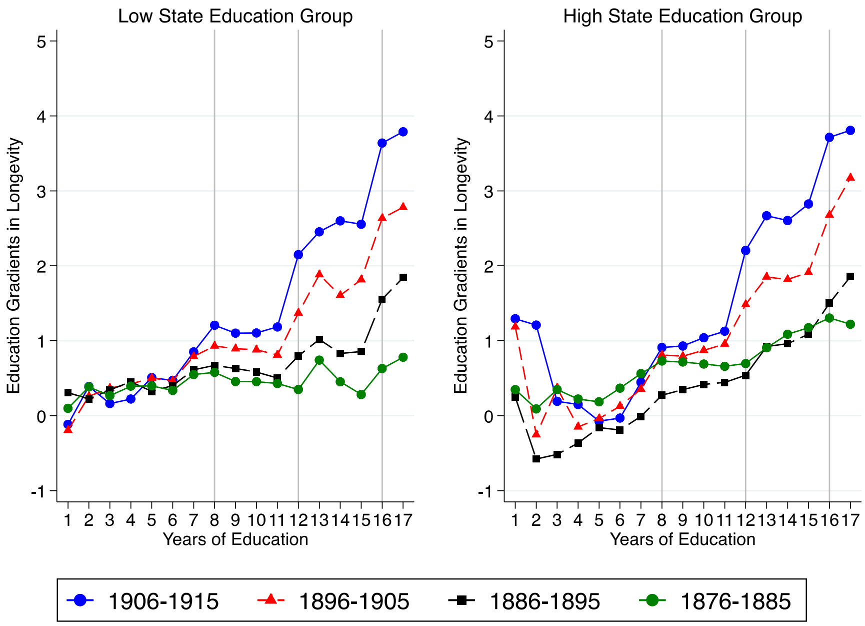


Fig. A11 Education gradients in longevity using CenSoc-Numident and Census-Tree Data. Analytic samples from CenSoc and Census-Tree data include whites born 1906-1915 in the 48 states and died 1988-2005. All models include a complete set of dummies for exact years of education (zero year of education as the reference) and adjust for state-of-birth fixed effects, year-of-birth fixed effects, and state-of-birth specific linear trends.

Panel A. For MALES



Panel B. For FEMALES

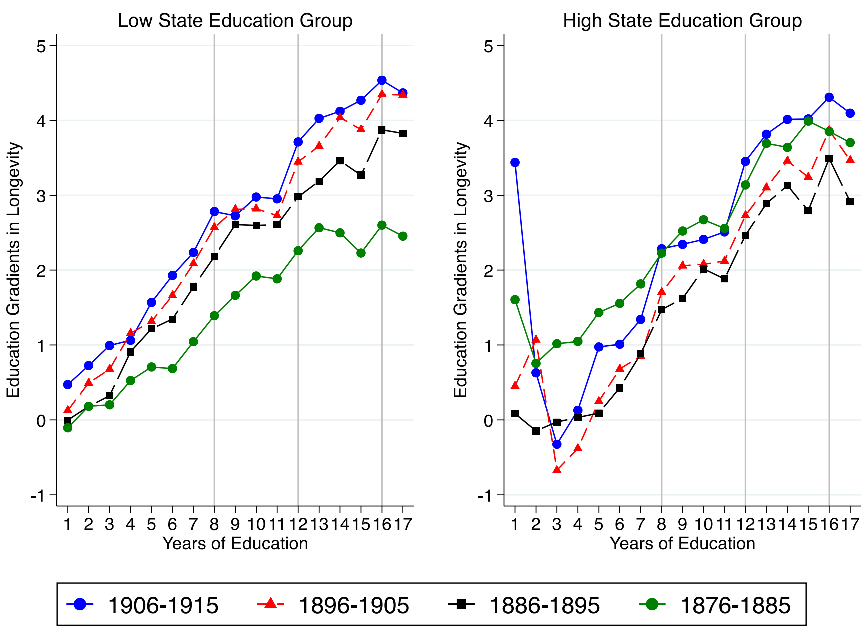


Fig. A12 Non-parametric estimates of the education-longevity relationship by gender, and birth cohort, and state-level education. The figure reports the coefficients from a regression of age at death on dummies for each single year of school, controlling for state of birth dummies, year of birth dummies and state-of-birth specific linear trend. The excluded category is 0 years of school. The 1906-1915/1986-1905/1886-1895/1876-1885 samples includes only whites born in the 48 states of the US who survived to age 65. States with above-median average education are included in the high state education group.