NBER WORKING PAPER SERIES

INTERGENERATIONAL CORRELATIONS IN LONGEVITY

Sandra E. Black
Neil Duzett
Adriana Lleras-Muney
Nolan G. Pope
Joseph Price

Working Paper 31034
http://www.nber.org/papers/w31034

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2023

This project was funded 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). Black received funding from the Research Council of Norway through its Centres of Excellence Scheme, FAIR project No 262675. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Intergenerational Correlations in Longevity
Sandra E. Black, Neil Duzett, Adriana Lleras-Muney, Nolan G. Pope, and Joseph Price
NBER Working Paper No. 31034
March 2023
JEL No. I1,I30,J6

ABSTRACT

While there is substantial research on the intergenerational persistence of economic outcomes such as income and wealth, much less is known about intergenerational persistence in health. We examine the correlation in longevity (an overall measure of health) across generations using a unique dataset containing information about more than 26 million families obtained from the Family Search Family Tree. We find that the intergenerational correlation in longevity is 0.09 and rises to 0.14 if we consider the correlation between children and the average of their parents' longevity. This intergenerational persistence in longevity is much smaller than that of persistence in socio-economic status and lower than existing correlations in health. Moreover, this correlation remained low throughout the 19th and early 20th centuries despite dramatic changes in longevity and its determinants. We also document that the correlations in longevity and in education are largely independent of each other. These patterns are likely explained by the fact that stochastic factors play a large role in the determination of longevity, larger than for other outcomes.

Sandra E. Black
Department of Economics and
School of International and Public Affairs
Columbia University
420 West 118th Street
New York, NY 10027
and IZA and NBER
sblack@columbia.edu

Neil Duzett
Texas A&M
1507 Hillside Drive
College Station, Texa 77845
neilduzett@gmail.com

Adriana Lleras-Muney
Department of Economics
9373 Bunche Hall
UCLA
Los Angeles, CA 90095
and NBER
alleras@econ.ucla.edu

Nolan G. Pope
University of Maryland
3114 Tydings Hall
7343 Preinkert Dr
College Park, MD 20742
npope@umd.edu

Joseph Price
Department of Economics
Brigham Young University
2129 WVB
Provo, UT 84602
and NBER
joseph_price@byu.edu

I. Introduction

A large literature in economics has documented the substantial persistence in economic wellbeing across generations, focusing attention on outcomes such as education, income, and wealth.[1] Much less is known about the persistence of health across generations. Richer, more educated individuals are healthier and live substantially longer lives in the US and most other developed countries today.[2] Moreover, this inequality in health appears to be rising.[3] But what is less well known is whether these differences in health outcomes are transmitted from parents to children. Yet, as recent efforts in economics suggest, health is an essential determinant of well-being, possibly as important as income.[4] Ultimately it is the intergenerational correlation in wellbeing that we care about, so understanding the intergenerational persistence in the components of wellbeing—in addition to how much these components move together—is fundamental to good policy.

In this paper, we use information from over 26 million individuals living in the United States between 1900 to 1920, and match these individuals to their parents and grandparents to estimate the intergenerational correlations in longevity. Correlations in longevity have several advantages over intergenerational correlations in socio-economic measures: they can be computed for both men and women, are easily comparable across time and space, and do not require any adjustments for price levels or other factors such as age. This makes it easy to track these correlations across many generations and for a variety of subpopulations.

To better understand how correlations in longevity relate to correlations in measures of socio-economic status, we also compute sibling correlations in longevity and compare them to correlations in income, education and occupation in the same sample. This then provides insight into the relative importance of intergenerational persistence in health and a broader understanding of how resources, genes, culture and environment are transmitted from one generation to the next. Finally, we can see how these sibling correlations—in longevity, income, and education—are related, to better understand the determinants of the persistence in wellbeing.

[1] See Black and Devereux (2011), Stuhler (2018) for summaries of this research.
[2] See Cutler et al. (2011), Marmot (2015), Chetty et al. (2016), and Galama et al. (2018).
[3] For example, the health gaps between the more and the less educated have grown bigger in the US: see Meara et al. (2008), Hummer and Hernandez (2013), and Case and Deaton (2017), although some important methodological questions about these results remain (see Leive and Ruhm, 2020 and references therein). There is also research showing that the variance of longevity is rising (van Raalte et al. 2018).
[4] See Jones and Klenow (2016), and Becker, Philipson, and Soares (2005).

In order to estimate the intergenerational correlations in longevity, we combine United States census records with data from the wiki-style Family Tree created by FamilySearch that includes over 1.4 billion people—the largest collection of its kind. We start with a dataset of all individuals from the 1900, 1910, and 1920 US censuses who were born between 1880 and 1920. We match these individuals to the FamilySearch database and gather data about their lifespan and the lifespan of their parents, grandparents, siblings, and spouses. We focus on individuals who survive to at least age 25, which provides us a sample of 26.1 million individuals matched with their parents and other family members. This final matched sample is fairly representative of the *white* population in the US census as a whole. By merging our sample to the 1940 US Census, we are also able to examine correlations in longevity between siblings, and to compare them directly using the same sample with correlations in other outcomes such as education and income, that have been studied more extensively in the literature.

To provide intuition on what intergenerational correlations in longevity measure and how to compare them with sibling correlations, we build a simple data-generating model to provide insight into the mechanisms through which longevity may persist across generations. With this model, we derive tests to disentangle the importance of genes versus family and social environment. Incorporating correlations across siblings as well as persistence across generations, we generate hypotheses for testing the different mechanisms at work. Using these data and guided by our model, we estimate the *intergenerational persistence in longevity* (IGPL) using a variety of transformations of longevity, including levels, logs and percentiles.[5]

We report four main results.

First, the intergenerational correlation in longevity is low and has been low since the early 1800s. Regardless of our choice of functional form, our estimates of the IGPL for one parent and their child (e.g., father lifespan and child lifespan) is low, ranging from 0.05 to 0.09. When we relate the longevity of children to the average of both parents' longevity the estimates increase to 0.14 for both sons and daughters. These estimates are robust to the inclusion of controls including demographic characteristics, cohort, and location fixed effects. When we investigate heterogeneity in our estimates over time, we find the IGPL remained fairly stable across the 1880-1920 birth cohorts, although increased during the early part of this time period.

[5] In this paper, we use the term IGPL to refer to both the intergenerational correlation as well as the coefficient from regressing the longevity of children on that of parents. Our conclusions do not differ based on the functional form.

We confirm these findings using a separate dataset of 4.7 million individuals from a subset of an earlier version FamilySearch Family Tree that covers a larger set of birth cohorts (1830 through 1920).[6] Interestingly, intergenerational correlations in education (Hertz et al. 2007) and occupation (Song et al. 2020) also exhibit relatively little change over the last century.

Importantly, the intergenerational correlation in longevity is lower than correlations in SES measures or health. Our estimates of the IGPL are substantially lower than estimates of the intergenerational correlations in income, education or wealth. For example, intergenerational income elasticities in the literature typically range between 0.30 and 0.40 (Black and Deveraux 2011).[7] We confirm these findings in our sibling sample, which shows that correlations in education are higher than correlations in income, and correlations in income are higher than correlations in longevity. We find evidence to suggest that the correlation in longevity is low, and lower than for other outcomes, because there is a large stochastic component in the determination of lifespan.

Second, multiple generations have independent effects on child longevity. Largely due to data constraints, the literature on economic mobility has primarily focused on the relationships between parents and children and has ignored the potentially important role of extended families in many societies (Mare 2011). We examine multi-generational effects on longevity to identify whether grandparents have an influence beyond that of the parents, or whether parents are a sufficient statistic for the family's role in longevity. Similar to work by Ferrie et al. (2021) and Lindahl et al. (2015) on education, and income and Johnston et al. (2013) for mental health, we find a positive role of grandparents in the child's longevity, even after controlling for parents' longevity. This suggests that either grandparents' longevity directly affects grandchildren or there may be some measurement error for parents' underlying health, for which longevity is a proxy.

Third, the role of genetic factors in determining longevity is modest, whereas the role of environmental factors is significant. In our sample of siblings, we can identify twins, the gold standard for studying genetic influences. Our model predicts that twin correlations will be higher

[6] Like the data used in previous research it is difficult to establish the representativeness of this sample. (See Appendix Table A1.) We compare the results from this long time series to the results in our representative sample for overlapping cohorts and find rather small differences.

[7] Intergenerational elasticities for wealth are typically around 0.37 in the U.S. (Charles and Hurst 2003). Intergenerational correlations in education on the order of 0.5 in the U.S. and range from 0.2 to 0.7 across countries (Hertz et al. 2007)

than non-twin sibling correlations, as they experienced the exact same family environment at the same ages. Identical twins also share the same genetic structure. Thus, identical twin correlations put an upper bound on the role of genetic factors. When we look at same-sex twins as a proxy for identical twins to get a sense of the relative importance of genetics, we find larger correlations in longevity among twins relative to non-twin siblings. However, even among twins, these correlations are low (0.18), and much lower than the twin correlation in education (0.64). Altogether, the results suggest that the genetic contribution to longevity is small.

Fourth, correlations in education and correlations in longevity are independent of each other in time and space. We assess this by looking at how sibling correlations in education for different cohort correlate with sibling correlations in longevity for the same cohorts. We repeat this exercise by state of birth and look at the association in the sibling correlation in education and the sibling correlation in longevity. In other words, the times and places where parents have transmitted their education to their children are not the same as those where they have been able to transmit their longevity. These results are similar to very recent studies (Halliday et al. 2019, Fletcher et al. 2023) which use substantially smaller samples to investigate similar questions.

Outside of economics, there are a number of early papers that documented the intergenerational transmission of longevity, beginning as early as 1899 with the classic work by Beeton and Pearson (1899, 1901). This literature is mostly concerned with identifying the role of genetics in longevity. Compiling a variety of datasets, they document that, among those living to age 20, the intergenerational correlation in duration of life between father and son is between 0.12 and 0.14, and between brother and brother is about 0.26. It is striking that, despite the significantly smaller datasets (1000 father/son pairs) and the much earlier time period they consider, the magnitudes we find are quite similar, although our findings are lower for sibling correlations. More recent work also finds remarkably small variation over time and space across studies in this correlation. (See Appendix Table A1.) However, the conclusions of these studies are usually limited by their small sample sizes and the fact that they are typically done with convenience samples that are not known to represent any specific population. A notable exception is the recent paper by Kaplanis et al. (2018) which uses a large (N=130,000) but not clearly representative sample. Like us, it finds low correlations in longevity and no trend from

1650 to 1850.[8] Our study uses a sample of more than 26 million individuals whose representativeness can be assessed because they are derived from complete census records. We are also able to compare our findings to correlations in other outcomes to provide a more comprehensive picture of intergenerational persistence.

There is also a burgeoning literature in economics on the intergenerational transmission of health.[9] A key challenge in this literature is the construction of a single health indicator that can be properly compared across generations. Recent work using register data from Denmark by Andersen (2021) documents intergenerational correlations in health indices based on health conditions (around 0.1-0.15). Using U.S. panel data, recent papers by Halliday et al. (2019, 2021) on intergenerational transmission of self-reported health status or health indices find somewhat larger correlations (around 0.26).[10] They also find that health correlations are lower than SES correlations and mostly independent of each other. Our conclusions – using a much larger dataset on earlier cohorts and looking at longevity – are similar: mobility in longevity is higher than mobility in other SES measures in both the US and in Scandinavian countries. Moreover, we find that mobility in longevity is larger than mobility in health measures, suggesting that quality of life, like SES and health, is more strongly transmitted across generations than quantity of life.[11]

---

[8] Other notable exceptions are, Ruby et al. (2018) use data from Ancestry family trees to examine the role of assortative mating in the estimation of heritability of longevity and find a substantial role for matching by around life span-influencing factors—either genetic and/or environmental. Kerber et al. (2001) and Gavrilov and Gavrilova (2001) are two other recent studies with relatively large sample sizes (N=78,994 and N=20,000 respectively). A complete list of these studies along with a summary of their findings can be found in Appendix Table A1.

[9] Work by Currie and Moretti (2007) and Giuntella et al. (2022) use data from California and Florida, respectively, and show significant correlations between mother's and child's birth weight (around 0.2). Using the 1970 British Birth Cohort Study Johnston et al. (2013) document correlations in mental health around 0.163. Two studies using the NLSY and the NHIS find correlations in BMI between 0.3 and 0.4 (Classen 2010; Classen and Thompson 2016). Thompson (2014) uses data from the NHIS and reports large correlations in chronic conditions, only a modest part of which appears to be genetically determined (20-30%). Most recently, Kumar and Nahlen (2023) document that the correlation in anemia between moths and children in India is 0.26. See recent handbook chapter by Halliday (*forthcoming*) for an excellent summary of this work.

[10] Health increases and is overall excellent among young adults, but it then declines with age. As a result, the number of medical conditions individuals report or are diagnosed with increases steadily from age 20 onwards. As is the case with income, intergenerational correlations will be sensitive to *when* the health measure is observed. There are many imperfect ways to address this issue. Longevity is not subject to this problem.

[11]There is also important new work trying to disentangle the role of nature versus nurture in longevity. Recent work by Bjorkegren et al. (2022) uses data on adoptees in Sweden to decompose this intergenerational transmission into nature or nurture. They find that the intergenerational association in mortality can be fully attributed to pre-birth factors; the association between the life expectancy of the biological parents of the children given up for adoption is as strong as for the children raised by their biological parents. Closely related is work by Hjelmborg et al (2006) that uses Danish, Finnish, and Swedish twins born between 1870 and 1910 to identify the genetic influences. They find that genetic influences on lifespan are minimal prior to age 60 but increase thereafter.

II. Data

We exploit a rich source of lifespan data that is available on a public wiki-style genealogical platform called FamilySearch. This platform includes over 13 million registered users and has profiles for over 1.2 billion deceased individuals. Most of these profiles are created by people who are doing research on their ancestors, and who gathered information on various aspects of the individual's life. A typical profile includes dates and places of vital events (birth, marriage, and death); sources that are attached to the person's profile (vital records, censuses, etc.); and also links to the profiles of their immediate relatives (parents, siblings, spouses, and children). This platform provides an open edit format so that anyone can make changes to any profile. Individuals doing research on the same ancestor contribute to the same profile and have the ability to correct any errors that are made by others. The wiki nature of this platform has pros and cons but provides one of the largest collections of intergenerational lifespan data that is publicly available. Previous research with similar genealogical platforms has shown that the information on family trees is quite accurate when verified using genetic data or vital statistic records (Kaplanis et al. 2018).

We designed a query of the Family Tree that uses a base sample of individuals that we gathered from the full-count US censuses for 1900, 1910, and 1920.[12] We focused on everyone who was born between the years 1880 and 1920, which provides a base sample of 173.3 million person-year observations. Since someone born before 1900 could have appeared in all three census years, we are likely to have some individuals appear in multiple census records. We are able to find personal identifiers (PIDs) for 133.3 million person-year observations in our base sample. These person-year observations collapse down to 89.3 million unique individuals.

After we have identified the PID for an individual from our census record, we can then use the information that is available on the profile page for that individual. We find that 39.0 million of these unique individuals have a death date on their profile. We also gather the PIDs for

[12] FamilySearch does not currently allow researchers to access the full corpus of profiles on the Family Tree. Rather, there is a public API that makes it possible to query the tree and gather information from the profiles of specific individuals. These queries can be based on information from a person's name or vital events, but the number of profiles that match a specific query must be less than 200 in order to access the data. It is also possible to use the public API to determine if an individual listed in a census record is attached to a profile on FamilySearch. This second approach requires access to the FamilySearch version of the US census records since it requires knowing the unique FamilySearch identifier for the specific person-record observation.

the individual's parents and siblings and then collect information about the birth and death dates of those family members. Of the people for whom we have a death date, we have a death date for at least one parent for 77% of them and a death date for both parents for 67% of them. We focus our analysis on those for whom we observe death information for both parents, yielding a final sample of 26.1 million child-parent groupings.

a. Data Quality

An important issue is whether the longevity data derived from the FamilySearch Family Tree are accurate. We use the Social Security Administration (SSA) cohort life tables by year of birth and sex to assess this.[13] Compared to the SSA data, our sample greatly underestimates mortality rates for infants and young children. This occurs either because individuals who are born between censuses and die young do not appear in a census, or alternatively, because their deaths are not noted in the Family Tree, a common problem in genealogical data (Kaplanis et al. 2018; Hollingsworth, 1976). However, if we condition on individuals living to at least 25 years old, the distribution of lifespan in our data is much closer to the distribution in the SSA data. This is shown in the top panel of Figure 1, where we plot the raw lifespan histogram from our sample and for the SSA data, from age zero (left top panel), and from age 25 onwards (right top panel). Conditional on surviving to 25, Figure 1 shows that the distribution of the age at death is shifted right in our sample, so that the mean age at death is slightly higher in our sample, but the shape of the distribution is similar.

This is also clear from survival rates. Using the Kaplan-Meier estimate for survival analysis, we plot the survival functions of the 1900 and 1910 cohorts by gender for both the SSA data and for our sample data conditional on survival to age 25 (bottom panels of Figure 1). Our sample tends to have somewhat higher life expectancy than the underlying population as reported by the SSA. For the 1910 birth cohort, the expected age at death for males (females) in our data is 70 (76), whereas it is 68 (74) in the SSA data. The differences are smaller for the 1900 birth cohorts for whom the gap is about a year.[14]

[13] Available here https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf.
[14] Appendix Table A2 present the gender breakdown of our sample relative to the SSA sample by age for the 1900 and 1910 cohorts as well.

Thus, while the distribution of the age at death is quite similar conditional on living to age 25, our sample lives a bit longer. Part of this difference is likely due to the fact that our final sample underrepresents African American and immigrants, two groups who have experienced higher than average mortality rates (Hacker 2010, Fogel 1986). Another reason is that the SSA tables were limited to data from states that had Vital Registration Systems in place. They re-weight their sample in an effort to make it representative of the US, but the weighting might not be sufficient. Our data includes all states.[15] We discuss the representativeness of the sample below.

b. Sample selection.

There is a natural asymmetry in the lifespan distribution of our base sample and that of their parents that stems from the fact that individuals don't become parents until living to a particular age. As a result, we never observe what the correlation in lifespan would have been for individuals who didn't live long enough to have children. To address this asymmetry and to address the issue of missing deaths among children, we restrict our sample to individuals at least 25 years old. We impose this restriction on both our base sample and on their parents. As we show later, our conclusions are robust to this restriction.[16] We also remove from our sample anyone who has a lifespan greater than 110 since, for the cohorts we consider, longer lifespans are likely a result of measurement/reporting error.[17]

c. Summary Statistics

Table 1 presents summary statistics for our analysis sample which includes the 26.1 million parent-child pairs with lifespan information for the individual and for both parents. Our data includes information from the original census record, including race, place of birth, and place of residence. We also obtained the total number of siblings and the individual's birth order from the Family Tree. These summary statistics make clear that our sample, while large, is not

[15] The SSA does not provide tables by state, so we are unable to directly assess how the inclusion of only a subset of states along with weighting affects the SSA estimates, or how much this explains the differences with our estimates.
[16] We prefer to condition to age 25 for several reasons. First it is most consistent with the prior literature in the study of heritability of longevity and in studies of the intergenerational correlation of SES measures—almost all previous work conditions on surviving to adulthood. (See Appendix Table A1.) Second and more importantly, our data does not represent children well.
[17] Because there is some evidence in the histograms that there is age heaping at 100, we also check the robustness of our conclusions to excluding individuals with lifespans 100 or over.

representative: Nearly all individuals in our sample are white (99%) and very few are immigrants (1%), though 13% have an immigrant father and 10% have an immigrant mother.

To assess this, in Table 2, we compare the characteristics of our analysis sample and the people from our original base sample from the census records who were age 25 or older. The table confirms that our sample severely under-represents African Americans and immigrants.[18] We do, however, have good representation of women, who are typically under-studied in analysis of intergenerational correlations (Hollingsworth 1976). Our data is also skewed towards the Midwest and under-represents the North relative to the full census data. Finally, our data under-represents earlier birth cohorts, most likely because many of them lived before birth and death certificates were universally required by law. By re-weighting the data by cohort and census region we can somewhat correct our estimates, as shown in Column 2 of Table 2. Finally, the table shows that we are able to track a larger share of the 25+ population: we observe about 2/3 of the baseline census population over the age of 25. Thus, while sample selection remains an important concern in our study, it is substantially smaller than in previous research.

Panel A of Figure 2 shows the trends in cohort life expectancy at age 25 for cohorts born from 1880 up to 1920. It shows several patterns consistent with the previous literature. First there is a steady increase in life expectancy starting in 1890, concurrent with public health investments in clean water and sewer systems (Cutler and Miller 2005; Troesken 2004, Alsan and Goldin 2019). Later gains are also consistent with the eradication of infectious diseases such as hookworm and malaria in the US early in the 20th century (Bleakley 2007; 2010b) and the availability of treatment for infectious diseases (in the form of sulfa drugs and antibiotics) in the 1930s and later. Cohorts born in 1880 lived most of their adult lives in an environment where infectious diseases were the largest killer, while cohorts born in 1920s experienced a world where chronic diseases--in particular cardiovascular disease--killed the most adults (Cutler et al. 2006). Second women live substantially longer than men (about 6 years in our sample, see table 1), and this female advantage grew substantially throughout the period.[19] This occurred in part

[18] Previous research has had similar issues with having appropriate representation of these groups — as a result many previous studies have restricted their focus to native-born Whites (Ward 2021).
[19] See Preston and Wang (2006); Beltrán-Sánchez et al. (2015); Cullen et al. (2016); Goldin and Lleras-Muney (2019).

because for men, life expectancy was falling in the 1880s, as has been noted elsewhere and consistent with the observed decline in heights among US men (see Costa 2015 for a review).[20]

We can examine trends over a longer period of time by making use of a smaller sample that covers cohorts born from 1820 onwards. This sample covers all individuals in the Family Tree who had non-missing data on their birth and death date and at least one parent with complete information as of 2018 when a snapshot of the tree was made available. These data come from an older version of the Family Tree; the advantage of this sample is that we can go back in time much further. However, we cannot assess the representativeness of this convenience sample. Panel B of Figure 2 shows the life expectancy trends in these data. It shows similar patterns for the 1880-1920 cohorts with dramatic increases in average longevity particularly for women. This older data reveals some other fascinating patterns. First, the gains in cohort longevity started early--much before public health efforts to eliminate infectious disease--and would be consistent with an increase in nutrition as argued by Fogel (1986, 2004). Also noticeable in these longer time series is that they show female disadvantage (women living shorter lives than men) for cohorts prior to 1870. These trends are to be taken with caution, however, since women are more poorly represented in older samples and the overall extent of selection for the older cohorts is not well understood.

Overall, both data sets show very similar patterns for longevity across the cohorts we study, and these patterns are in line with the existing historical evidence, suggesting our data is of high quality. Before moving on to our analysis, it is worth noting that the fact that longevity has increased across cohorts has no immediate or mechanical implications for intergenerational correlations. If children live 20 years longer than the parents, but longer-lived parents always have longer-lived children, then the correlation in longevity across generations will remain stable. Conversely life expectancy could be stable, but the IGPL could change.

III. A Simple Model of the Determination of Longevity

[20] The reasons for this decline are poorly understood but it is thought to relate to the worse infectious disease environment and pollution associated with rapid urbanization during industrialization (Costa 2015, Cutler et al. 2006). Trends in life expectancy in the US are hotly debated because of data concerns. There are considerable differences in different time series estimates – see Hacker (2010) for a comprehensive discussion. There appears to be consensus that period life expectancy fell sometime in the 19th century and then started increasing later on, but there is no consensus on the exact timing of the reversal. Also, note that there are no cohort tables for the 19th century, and it is difficult to make predictions about cohort trends (like the ones shown here) based on period tables.

We now present a simple model to provide a framework for better understanding the intergenerational correlations in longevity and their relationship to sibling correlations. Previous research on the determinants of longevity suggests that longevity is a function of genetics, environmental factors (rain, temperature, the quality of the air, the availability and quality of the food, access to and quality of life saving technologies, the quality of social interactions, etc.), gender, socioeconomic status (education, income and occupation), and random components. (See Cutler et al. 2006, van den Berg et al. 2017.) We start with a very basic model that treats these factors as linearly additive in the determination of longevity.[21]

We denote L^g_{ijsc} as the lifespan (or age at death) of an individual *i,* with gender *g*, from family *j*, living in place *s*, and in cohort *c.* Suppose that the individual's lifespan is determined uniquely by their genes (G^g_{ijsc}), environmental factors that are common to all individuals living in the same place and cohort (α_{sc}), a gender-specific factor common to each gender (γ_i^g), family's social economic status (θ_j^{SES}), and random factors (e_i). Let G^f_{jsc-1} and G^m_{jsc-1} denote the genes of the female parent (mother) and the male parent (father).[22] Each child in family *j* receive these genes along with variation in the genes they draw from each parent (η^f_{ijc} and η^m_{ijc}).[23] Children receive half of their genes from each parent: $G^g_{ijsc} = \frac{1}{2}\left[G^f_{jsc-1} + \eta^f_{ijc} + G^m_{jsc-1} + \eta^m_{ijc}\right]$. We assume G^f_{jsc-1} and G^m_{jsc-1} are random variables with an unknown variance-covariance matrix reflecting the extent of assortative mating that is gene-based. The variables η^f_{ijc} and η^m_{ijc}, are assumed to be distributed with mean zero and an unknown variance-covariance matrix, where the covariance is assumed to be zero. In this model then we have that for child *i* in family *j,* lifespan is given by:

$$L^g_{ijsc} = \delta\frac{1}{2}\left[G^f_{jsc-1} + \eta^f_{ijc} + G^m_{jsc-1} + \eta^m_{ijc}\right] + \alpha_{sc} + \gamma_i^g + \theta_j^{SES} + e_i \qquad (1)$$

where e_i is a random variable, assumed to be independent across individuals, and from *G*, α_{sc}, γ_i^g, and θ_j^{SES}. The parameter δ represents the effect of genes on lifespan, where the genes are

[21] Throughout, we use the terms longevity, lifespan and age at death interchangeably, although note that in some fields, longevity refers to whether individuals live a particularly long life and is studied separately from lifespan (van den Berg et al. 2017).
[22] Gender captures biological, hormonal and social effects.
[23] Here *c-1* represents the parents' cohort.

assumed to affect lifespan through a linear index, such as a polygenic score.[24]

We assume that the common environmental component (α_{sc}), the gender-specific component (γ_i^g), and social economic status (θ_j^{SES}) are independent of the genes *G* and of other random individual specific shocks.[25] In addition to assuming linearity in genes and environment, Equation (1) assumes that there are no interactions between genes and environment which have been shown to exist.[26] We are also ruling out interactions between gender and the environment.[27] We discuss the implications of these assumptions later, and we relax them in estimation. But this simple model provides a useful baseline.[28] This data-generating model has strong empirical implications, summarized here in proposition 1 (the Appendix outlines the covariances in longevity from this model).

Proposition 1: *For a closed population (where the joint distribution of G is constant) living in a stable environment ($\alpha_{sc} = \alpha_s \; \forall c$), the following will be true:*

a. The expected lifespan of fathers (mothers) and sons (daughters) is identical, and so is the variance. (However, in general, we are not in a stable environment, with changes in public health policies and medical technologies, suggesting the importance of controlling for cohort effects in the basic regression.)

---

[24] In this framework *G*, or more precisely the function $\delta_2^1\left[G_{jsc-1}^f + \eta_{ijc}^f + G_{jsc-1}^m + \eta_{ijc}^m\right]$, can be viewed as a polygenic score that combines all the genes that are known to determine longevity. A polygenic score is a weighted sum of different genes, where the weights have been estimated typically by a GWAS study. For example, Deelen et al. (2019) and Timmers et al. (2019) identify genes that affect longevity. Timmers et al. construct a polygenic score to predict longevity using the identified genes.

[25] Evolutionary biology shows that over long periods of time environmental conditions will affect the set of genes that survive in the population. Thus, the assumption that the genes and the environment are uncorrelated is likely incorrect. This has been noted before e.g. by Manski (2011). Recent papers in genetics have demonstrated that genes are indeed correlated with environments. For example, Belsky et al. (2016) show that children with higher polygenic scores for education are born into more advantaged homes.

[26] For extensive discussions of these interaction effects and how they affect studies of the heritability of lifespan see the review by van den Berg et al (2017).

[27] This is also unrealistic. Life expectancy for women is higher and has grown more than that of men in the 20th century, which suggests environmental changes have favored females, see Goldin and Lleras-Muney (2019).

[28] This model has some important limitations. The model is mechanistic. The only reasons why life expectancy and intergenerational correlations in lifespan change is because environmental conditions change exogenously across space and time. This model does not allow individuals to invest in their health and longevity or to optimally chose the locations that would maximize their longevity. We also exclude the possibility that the longevity of the parents has a direct impact on the longevity of their children, conditional on the genetic material that parents give to their children, which is very unlikely to be true. In addition, the model does not allow for a person's social economic status (which is also highly heritable) to affect the longevity of individuals (which empirical evidence shows matters) – we assume only the family's SES matters.

b. *The simple bivariate intergenerational correlation in longevity between parents and children captures both genetic components and other factors,* namely: how much genetic assortative mating there is on the part of the parents, the extent to which parents and children share the same environment, the effects of SES on longevity, and the variance of the gender component. Thus, this correlation does not uniquely describe the extent to which genes affect longevity, as has been noted previously (e.g. see review by van den Berg et al. 2017).

c. The intergenerational covariance between father (mother) and son (daughter)'s lifespan is equal to the male (female) sibling covariance.

d. *The covariance between twins' lifespans is greater than the covariance between the lifespans of siblings of the same gender*. This occurs because their genes are the same.

e. The male (female) sibling covariance is larger than the opposite-gender sibling covariance, if $V(\gamma_i^f) \neq 0$ *and* $V(\gamma_i^m) \neq 0$.

f. The father-son (mother-daughter) covariance is larger than the father-daughter (mother-son) covariance, if $V(\gamma_i^f) \neq 0$ *and* $V(\gamma_i^m) \neq 0$.

g. The intergenerational covariance between a child and the average lifespan of the parents is the same as the intergenerational covariance between a child and either of the parents. However, the variance of the average lifespan is approximately half the variance of a single parent's lifespan, if the $COV(G_{jsc-1}^f, G_{jsc-1}^m)$ *is small relative to* $V(G_{jsc-1}^f)$ *and* $V(G_{jsc-1}^m)$.

h. The intergenerational covariance between a paternal grandfather (grandmother) and grandson (granddaughter)'s lifespan is less than the covariance between father (mother) and son (daughter)'s lifespan, if the $COV(G_{jsc-2}^{paternal\ m}, G_{jsc-2}^{maternal\ m})$ *and* $COV(G_{jsc-2}^{paternal\ m}, G_{jsc-2}^{maternal\ f})$ *is less than* $COV(G_{jsc-2}^{paternal\ m}, G_{jsc-2}^{paternal\ f})$ *and less than* $V(G_{jsc-2}^{paternal\ m})$.[29]

A few observations about these statements are useful. If the environmental factors that

[29] In other words, the cross (maternal to paternal sides) grandparent genetic "assortative matching" has to be less than the actual spousal genetic assortative mating of grandparent and less than the variance of the grandparent's genetics. Both of these conditions are very likely to be true.

determine lifespan are not changing, the variance of lifespan is constant for each gender across cohorts (Prop 1a). This implies that all statements about covariances also hold for bivariate regression coefficients (except for part g, in which the regression coefficient between a child and the average lifespan of the parents will be approximately twice as large as the intergenerational covariance between a child and either of the parents). We can then summarize the predictions of the model as stating that in a stable environment for males:

$$\beta_{twins} > \beta_{brothers} = \beta_{father-son} > \beta_{mother-son} > \beta_{father-mother} \quad (2)$$

where the β is the coefficient from a regression of i's lifespan on the lifespan of their parent or sibling. The same will hold for females. However, a priori it is unclear whether father-son covariances will exceed mother-daughter covariances.[30] Similarly, it is also unclear whether brother covariances will exceed sister covariances.[31]

The closed population assumption assumes there is no significant in- or out-migration of genetically diverse individuals.[32] This assumption might not hold in places like the US during the 19th and early 20th century. Migration alone however does not necessarily lead to a violation of the assumption, so long as migrants are drawn from a similar genetic pool.[33]

IV. Empirical Approach and Results

a. Empirical specification

Our main specification relates the lifespan of the child (L_i^C) to the lifespan of the parent (L_i^P). We estimate the following equation:

[30] This will depend on whether the variance in the female-specific component is larger or smaller than variance in the male-specific component (i.e. whether $V(\gamma_i^f)$ or $V(\gamma_i^m)$ is larger).

[31] The prediction that $\beta_{brothers} = \beta_{father-son}$is in contrast to other models (see Solon 1999 for one example) that predict that the sibling correlation will be different from that of intergenerational persistence. This is a result of our model specification, where we assume that parents and children share 50% of their genes (as do non-identical twin siblings) and that parents and children grow up in the same environment (as do siblings). In more nuanced sibling models, there is a family background component that is not shared between siblings (due, for example, to variation in the age of siblings), and in intergenerational models, there is a family background component that is not shared between parents and children.

[32] The assumption that the distribution of genes in the population is constant further rules out large-scale genetic modifications such as those that would be due to, for example, massive exposure to nuclear waste.

[33] This is more likely to hold if migrants come from the same countries every generation. If the distribution of genes is constant, and the environment is stable, then in this model migrants help identify the effects of different spatial environmental conditions on lifespan (α_s). This is a broader statement that is consistent with the often-used empirical approach that uses identical twins reared in different environments to assess the impact of the environment separately from the genetic influences. As noted earlier, this decomposition is only possible under the strong assumption that there are no important gene-environment interactions.

$$L_i^C = \beta_0 + \beta_1 L_i^P + X\beta_3 + \epsilon_i \qquad (1)$$

where L_i, our main variable of interest, is longevity in years for individual i and L_i^P is the longevity of one of the parents, conditional on living to age 25 for both.[34] We also allow L_i^P to refer to the average of the parent's longevity, as is frequently done in the heritability literature. X refers to a set of control variables; our main estimates include a parsimonious set of controls, including indicators for the cohort of the child and the cohort of the parents to account for secular trends in lifespan. In some specifications, we also include state of birth fixed effects to proxy for the environmental factors described in the model, as well as controls for race and immigrant status.[35] The standard errors are clustered at the family level since our sample can include multiple children from the same parents. As the model is agnostic about specific functional form, we present results in levels, logs, and percentiles. We further investigate whether this functional form is correctly specified as linear.

The coefficient of interest is β_1. In levels, β_1 represents the average increase in longevity associated with a one-year increase in the longevity of the parent. The intergenerational correlation in longevity can then be expressed as $\gamma = \frac{\sigma_p}{\sigma_c} * \beta_1$, where σ_p represents the standard deviation in longevity for the parents and σ_c is the standard deviation for the children. When the model is estimated in logs, β_1 represents the intergenerational elasticity of longevity.[36] When L is converted into a percentile, β_1 represents the Spearman correlation in longevity. In these specifications, regardless of functional form, the coefficient on longevity, β_1, will incorporate the influence of both parental genetics as well as socio-economic status. In later specifications, we will try to disentangle the role of genes and environment.

As is typical in this literature, we are estimating associations and not the causal effect of exogenously changing parental longevity. As our model suggests, there are a variety of factors that can influence children's longevity beyond parental longevity—including environmental factors as well as the state of current and past medical technology and public health policies. However, we will show results with and without covariates to assess whether the relationship appears to be subject to substantial omitted variable bias. We find the relationship remarkably

[34] We later test the sensitivity of our conclusions to this restriction.
[35] From the model, this will correct for the non-stationary aspect of longevity.
[36] Note that in a stable environment these standard deviations are identical and so the correlation in logs and the elasticity would be expected to be the same.

robust to the inclusion of additional controls.[37]

b. Main results

We start by estimating the relationship between the lifespans of children and their parents and then test the robustness of this relationship to the inclusion of various controls (Table 3). We find that the raw coefficient from a regression of the lifespan of a son on the lifespan of his father is 0.09 (Column 1), while that for a son and his mother is 0.06. When we relate the lifespan of a son to that of the average of his parents' lifespan, we see the coefficient is significantly higher, at 0.14.[38] Similarly, when we look at daughters, we see that the daughter/father relationship is 0.07 and daughter/mother is 0.08. When we relate the lifespan of a daughter to that of the average of her parents' lifespan, the coefficient is 0.15.

Although the time series show that lifespan was increasing among our cohorts, we obtain nearly identical coefficients when we include parent and child birth-year fixed effects (Column 2). This is our preferred specification. When we include state-of birth fixed effects for both parents and children (Column 3), again, the results are unchanged. Finally, in the last column (Column 4), we add additional controls for race (Black, White, Other) and birth order. While these characteristics have significant effects on child longevity, the coefficients on parental lifespan are unchanged.[39]

We find that father-son coefficient (0.09) is larger than mother-son coefficient (0.06), consistent with the predictions of the model. But the mother-daughter coefficient (0.07) is nearly identical to the father-daughter coefficient (0.07) once controls are included. It is not clear why this is the case. The coefficients for daughters tend to be smaller than the coefficients for sons, but these gender differences vanish when we use the average of parents as a regressor: the son/parent or daughter-parent coefficient is approximately 0.14.

[37] In a later section, we also examine the inclusion of individual-level socioeconomic controls as an effort to identify the mechanisms underlying the intergenerational persistence.

[38] This is consistent with both the fact that there is likely measurement error in our underlying variable of interest ("parent health") as well as the fact that there is independent information contained in each parent's longevity, as discussed later.

[39] We also find that the correlation in lifespan between the fathers and mothers is positive but modest, only 0.04 (last row of Table 3). Thus, there is some amount of positive assortative mating based on health, but it is small compared to, for example, correlations in education across spouses today, which are frequently higher than 0.5. (For example, the correlation in education across spouses in the PSID is 0.6 according to Oreffice and Quintana-Domeque (2010). See Blossfeld (2009) for a review.) This result is consistent with the findings in Domingue et al. (2014) who find that assortative mating based on genetic similarity is much smaller than assortative mating based on education.

In Table 4, we use the same specification as Column 2 in Table 3 (with cohort fixed effects), but vary the functional form, starting with the levels specification (Column 1), the levels specification reweighted using the census data to better represent the population (Column 2), a rank-rank specification (Column 3), and the log-log specification (Column 4).

Interestingly, the patterns look quite similar to those in the previous table. For sons, we find the coefficient on father's lifespan is larger than the coefficient on mother's lifespan and both are smaller than when we average the two parent's lifespans together. For daughters, the coefficients on fathers and mothers are closer, with the ranking varying depending on the specification. Again, the coefficients are larger and identical to the son's coefficients when we average the two parents' lifespan. For both sons and daughters, the weighted estimates (Column 2) are somewhat lower than the unweighted estimates. Since the weights adjust for cohort and spatial differences, this suggest that there are differences in these coefficients in time and space, an issue to which we return below.

We get similar estimates when we use percentiles of lifespan (Column 3) or a log-log specification (Column 4) although the log-log specification yields the smallest coefficients among all specifications. Appendix Table A3 shows raw correlations instead of regression coefficients, since these are often reported in other studies. Raw correlations in lifespan (or its transformations) are even lower than regression coefficients. For example, the raw father-son correlation is 0.08 (instead of 0.09) and the son-parent correlation is only 0.10 instead of 0.14.

To check that our linear specification is appropriate, we visually plot the relationship between the lifespans of parents and children (Figure 3). We do this both for the lifespan and the lifespan percentile/rank. We find that, for lifespan, the relationship is fairly flat for those whose parents died before age 40 and then becomes steeper and remains roughly linear. We hypothesize that the flat relationship for parents who died young could be due to many of these early deaths being accidental deaths, which result in the parent's lifespan having less meaningful information about the underlying health characteristics of the parent. For rank, the relationship is linear at almost all ranks except for the very bottom and the top 5%.

Another way to assess the extent of persistence and whether the linear specification misses important patterns is to estimate transition matrices, which show the probability that a child born to a parent in a given quintile ends up in that quintile, or in a lower or higher one. Appendix Tables A4 and A5 show that the transition matrix for both women and men is close to

what we would expect if there were close to perfect mobility: the diagonal elements are somewhat larger than 20% but not much, and the off-diagonal elements are generally a bit lower than 20% but not substantially. The only entry that appears to substantively deviate from perfect mobility is the probability of living a long life (being in the top quintile), which is 25% for sons (daughters) of fathers (mothers) in the highest quintile of lifespan. This is consistent with findings in the literature but suggests these effects are small.[40]

These results suggest there is significant upward and downward mobility, much more so than for income today. Indeed, the son (daughter) of a father (mother) who was in the bottom quintile of lifespan has a 17.5% chance of being in the top quintile of the lifespan distribution. By contrast among cohorts born in 1970s, the probability that a child born to the poorest 25% of parents ends up in the top 25% of the income distribution is only 10% (Chetty et al 2014). Conversely, the son (daughter) of a long-lived father (mother) (in the top lifespan quintile) has a 16.5% chance of being in the bottom quintile of the lifespan distribution. This is substantially larger than downward mobility in income: the probability that a child born to the richest parents ends up in the bottom 25% of the income distribution is only 9% Isaacs et al. (2008).

Finally, up to this point we have conditioned on the child being alive at age 25. Prior research has used different cutoffs, ranging from age 15 to age 65. (See Appendix Table A1). In addition, many studies focused on the role of genes have argued that, to identify genetic effects, it is more appropriate to condition on living to very old ages, as younger deaths are more likely due to accidents. Importantly, Appendix Figure A1 shows these correlations are remarkably robust to the choice of cutoff ages and they remain low throughout. These results suggests that genetics contributions to longevity are small an issue we return to later.

c. Grandparents

While longevity is clearly correlated across generations, is this correlation limited to one generation, or does this correlation persist across multiple generations? Recent research on intergenerational education and income correlations have documented an independent effect of

[40] For example, Perls et al. (2002) and Schoenmaker et al. (2006) show that the survival rates of siblings of long-lived individuals is higher than that of the population.

grandparents' outcomes even after controlling for parent's outcomes, although evidence suggests this may be driven by measurement error in the underlying parameter.[41]

Our model predicts that correlations between grandparents and grandchildren should be smaller than correlations between parents and children. Table 5 presents the results when we examine all three generations, using the subsample for which we know the lifespan of all grandparents. Columns 1 and 2 present the results from bivariate regressions of children's lifespan on that of their parents or grandparents, separately for sons and daughters, controlling only for birth year effects for the child and the relevant parent/grandparent(s). Each cell reports the coefficients from a different regression.

The bivariate regressions show that although this three-generation sample is much smaller, the father-son IGPL is very similar (0.099) to the overall IGPL in the large sample (0.090) and the daughter-mother coefficients are also very close (0.075 vs 0.074).[42] Thus the sample does not appear to be particularly selected. We also find that parents' longevity is more highly correlated with children's longevity than grandparents' longevity, with magnitudes about a third as large for grandparents relative to parents. This result is consistent with the predictions of the model.

In Columns 3 and 4, we include both parents' lifespan as well as grandparents' lifespan in the same regression (so each column represents a single regression). Including both mother's and father's longevity at the same time only minimally reduces the coefficient on either one, suggesting a role for both mothers and fathers separately. These independent contributions of mother's and father's longevity is not surprising given the low correlation between father's and mother's lifespan of 0.04 in Table 4.

The inclusion of grandparents' lifespan reduces the effects of parents' lifespan by a very small magnitude. Consistent with the existing research, and contrary to the predictions of the model, even with the inclusion of parents' lifespan, grandparents' lifespan has an independent effect. This could reflect either that long-lived grandparents have a direct impact on grandchildren's longevity, for example by caring for them in childhood (as shown for example by Luo et al. 2012), or that there is measurement error in our proxies – an issue to which we

[41] There has been recent work documenting the importance of grandparents and other generations in intergenerational correlations of outcomes such as education. See Lindahl et al. (2015) and Ferrie et al. (2021).
[42] The larger difference among women is likely because selection into the sample is likely greater among women. Historical records of women are known to be less frequently available and more subject to error than those of men, see Hollingsworth (1976).

return. For sons, fathers and grandfathers appear to have a larger influence than mothers and grandmothers. For daughters, grandmothers and grandfathers have a similar influence.

d. Heterogeneity across time

So far, we have pooled all cohorts and simply controlled for parents' and child's year of birth. However, there are a number of reasons we might expect the intergenerational correlation in longevity to change over time. First, we saw that life expectancy changed dramatically over our period of study; the public health and technological changes that allowed for life expectancy to increase could lead to a lower correlation between parents and children, if for example children from low SES backgrounds, with likely low predicted longevity, benefit more from these changes.[43] On the other hand, as incomes and income inequality rose, it is possible that IGPs in longevity and other SES measures would become stronger.[44]

Figure 4 shows the IGPL by year of birth of the child and by gender (Panel a). For men, we find that the IGPL increases from around 0.06 to 0.10 over the first two decades of the sample and then plateaus from 1900 to 1920. Women see a similar increase over the first two decades, but then see a decline over the last two decades to roughly 0.07. The IGPL between children and both parents increases from around 0.10 to 0.15 during the first 20 years and then plateaus. While somewhat muted, a similar pattern appears in the data when using a rank-rank specification (panel b). While the IGPL remains low over this entire period, we see a meaningful rise from 1880 to 1900. Both figures show that the patterns are very similar in the data that is matched from the census (our main sample), or the data that simply comes from the family trees.

We also show the patterns over a much longer time period (for cohorts born 1830-1920) using the convenience snapshot we obtain of the Family Tree. Figure 5 shows that the father-son IGPL increased over second half of the 19th century, from a low of roughly 0.05 in the mid 1800s to about 0.1 in 10. However, for the mother-daughter correlations the IGPL remains constant

[43] For example, as argued by Troesken (2004), Anderson et al. (2021) find that city-wide chlorination lowered black but not white infant mortality rates. Other innovations have been biased towards high SES individuals. For example, Preston and Haines (1991) show that the infant mortality rate was lower among literate mothers in the early 20th century. Theoretically it is unclear whether health innovations increase or decrease health inequality.

[44] It is also possible that the data on longevity for earlier time periods are of lower quality: birth and death certificates were not required or used in many states before the 1930s, and literacy levels were low. In this case the correlations would be lower than expected in the distant past due to measurement error, but increase in more recent periods as the quality of reporting rose.

more constant throughout with less growth throughout the 1800s, although the data is noisy. Again, the results are similar in levels (panel a) and ranks (panel b).

While these results suggest some heterogeneity over time, the IGPL remains low throughout this time period.[45] Even with the vast changes in environmental conditions that cohorts faced from 1880 to 1920, the IGPL was always at or below 0.10, and reaching a high of 0.15 if we look at correlations with the average of both parents for the more recently born cohorts. By contrast the rank-rank correlation in income or occupation-based income has remained around 0.3 for all cohorts of men born since 1880 (Chetty et al. 2014; Song et al. 2020). Interestingly however the correlations in income were lower and grew for cohorts born before 1860 which is similar to what we find (Song et al. 2020). This suggests that the cohorts born in the early 19th century experienced substantial mobility in both health and income. Song et al. (2020) find that the larger income mobility in the 19th century was due to the move from agriculture to manufacturing. This move may have allowed for more mobility in both incomes and health than the economic environments thereafter.

e. Sibling Correlations

While our primary interest is in estimating intergenerational persistence in longevity, examining sibling correlations provides complementary information about the combined role of genetics and environment. As our model describes, the basic components are the same – genes, environment, socioeconomic status and gender will all factor into the sibling correlations. A general challenge when comparing correlations across outcomes using estimates from other studies is that one cannot isolate whether observed differences are due to true differences in correlations or whether differences result from using different samples. In order to directly compare the correlation of lifespan and the correlation of other economic outcomes *in the same sample*, we estimate sibling correlations. Since we are focusing on only one generation, we can examine a number of different outcomes—including education, income, and longevity—on a large sample of individuals.

To conduct the analysis, we construct a sample of sibling pairs whose education,

[45] As noted in the introduction these findings are consistent with Kaplanis et al. (2018)'s findings that the IGCL exhibits no trend over a long period of time. Their estimates, although derived from a smaller sample, cover the children of parents born from1650 to 1850.

occupation, and income are observed in the 1940 census.[46] Appendix Table A6 shows that our sibling sample is observably very similar to the full analysis sample. We divide our analysis here to compare brothers with brothers, sisters with sisters, and sisters with brothers. To estimate these correlations, we replace the parents' longevity with that of the sibling (or, in the case of multiple siblings, with the average of the siblings).

In Table 6 we estimate sibling correlations in longevity, education, and earnings (both individual and household) using a constant sample (columns 1 - 4). Because many women do not have positive earnings in the 1940 census, we also consider a larger sample restricted to those with both education and longevity as a robustness check (Columns 5 and 6).

Several conclusions emerge. First, the correlations in longevity among siblings are larger than our parent-child intergenerational correlations: for example, the brother-brother correlation is 0.13 instead of 0.08 (the father-son correlation shown in the last column for this sample). This is a deviation from the model, which predicts these would be identical, although would be consistent with other models with slightly more nuanced treatment of environment.[47] In the model, this would happen if, for example, the environment was more similar among brothers than between parents and children.

Second, gender plays a similar role among siblings, as described in the model. Similar to the child-parent correlations, we see a stronger correlation between male siblings' lifespans (0.13) than female siblings' lifespans (0.11). And we see a rather weak correlation between brothers and sisters (0.04).

Third, the correlations in SES we document mirror those in the existing literature, suggesting our sample may be broadly representative. The coefficient on education is 0.55 for brothers, 0.60 for sisters, and 0.53 for mixed sex siblings. When we look at correlations (Appendix Table A7), we get correlations of 0.55 for brother, 0.59 of sisters, and 0.53 for mixed sex siblings. This is quite consistent with the literature. For example, Solon et al. (1991) estimates that the sibling correlation in education is 0.45 for male cohorts born in 1944-1958 and

[46] We construct our sample of siblings by using people in our data who have the same two parents and for whom both siblings are attached to the 1940 census on their profile on the Family Tree. We do not do a similar exercise for parents and children because it would require reducing our sample to those parents who survived to 1940, the first year when education and income are observed.

[47] See Bjorklund and Jantti (2012) for only one example. Using data from Sweden, they estimate brother correlation in schooling of 0.46 and father-son correlations of 0.39; similarly, the brother correlations in earnings is 0.24 while father-son correlations are 0.14.

Bjorklund et al (2002) estimates a sibling correlation of 0.43 for cohorts born in 1951-1957. Our estimates are somewhat larger for women, Solon et al. (1991) estimate a 0.28 correlation for the 1951-1958 birth cohorts and Mazumder (2008) estimates a 0.34 correlation for the 1947-1955 cohort.

When we look at income, we see a similar picture. We estimate a coefficient of 0.25 for males with a correlation of 0.26, which is comparable to those estimated in the literature, which generally ranges from 0.1 to 0.4 for those using single-year earnings.[48] We also show coefficients for the relationship in household (rather than individual) income to overcome the fact that most white women in this period did not work and thus would have no income to report.[49] Estimates using household income are larger than those for individual income. Also, the coefficients for sisters (0.36) or brother-sister pairs (0.33) are now similar to coefficients for household incomes among brothers (0.35).

Fourth, using a fixed sample of individuals, we see much larger coefficients (and correlations) for education or income than for longevity. The coefficients for household income are at nearly three times as large as those for longevity: they range from 0.33 for brother-sisters to 0.36 among sisters. The correlations for education are even larger ranging from 0.53 for brother-sister to 0.60 for sisters, more than four times larger than the correlations for longevity.

In sum, we find that, consistent with intergenerational lifespan correlations, sibling lifespan correlations are also low: they range between 0.11 and 0.13, and they are substantially smaller than correlations in all other SES measures. Sibling correlations exceed intergenerational correlations; this is likely due to sibling shared environmental conditions that are not shared between parent and child, at least not at the same age.

f. Mechanisms: Why is longevity persistent across generations?

As discussed in our model, the intergenerational correlation is a function of gender, socio-economic status, genes and environment; however, we do not know the relative importance of these factors. While we are unable to identify specific causal mechanisms, we use a variety of means to try to disentangle genetic and environmental components.

[48] See Solon (1999).

[49] The brother-sister correlations in income are negative. We hypothesize this is because parents tended to marry off their daughters and provide their sons with an occupation. When we look at household income the correlations are in fact very similar.

a. Controls for Socioeconomic Status

As a first pass, we examine how sensitive our estimates are to controls for socioeconomic status. If socioeconomic status is an important component in parents' longevity that is also important for children's longevity, its inclusion in the regression should reduce the coefficient on parent's longevity. Unfortunately, we do not observe any measures of parental SES (beyond longevity). We are therefore limited to using a limited set of controls for the child's SES: their education, income and occupation. These results are presented in Table 7. For both sons and daughters, these SES controls do not substantially affect the coefficients on parental lifespan. The coefficient on education is positive and significant, consistent with prior research for these cohorts, but its inclusion does not materially affect the coefficient on lifespan.[50] The coefficients on income and occupation appear to have positive and significant effects on female longevity but not that of men, consistent with the controversies surrounding the role of income on health (Cutler et al. 2006). Importantly, adding these controls has no effect on the parental longevity coefficients for either males or females, suggesting a limited role for socio-economic status as the main explanation for the persistence in longevity.

b. Sibling Types

Another way to try to disentangle genes from environment in the intergenerational correlation in longevity is to take advantage of having both twins and non-twins in our sample, who all have varying degrees of genetic connections but who grow up in the same family.

In our simple model, the only reason identical twins have different lifespans is because of their individual-specific random shocks e_i; we would thus expect very high correlations in longevity across identical twin pairs. If we assume that twins are growing up in the same environment, the difference between the correlations among siblings and among identical twins provides insight into the role of genetics. Among fraternal twins, in addition to the random shocks e_i, that identical twins experience, there is also variation in the genes they received from their parents (η_{ijc}^{f} and η_{ijc}^{m}), and they may be of different genders. Other characteristics, such as

[50] The coefficient on education is similar to the coefficient reported in Lleras-Muney et al. (2022) who find a coefficient of education of 0.4 (in a regression of longevity on education conditional on being alive at age 35) for the 1906-1915 cohort, and smaller coefficients for older cohorts.

environment and socioeconomic status, should be very similar. Siblings' lifespans may also differ if they are raised in different locations and therefore exposed to different environments. They also differ because siblings are by definition not born in the same year and thus subject to cohort effects. The differences in these coefficients can thus be used to make inferences about the relative importance of genes and environment in terms of longevity.

We identify twins as siblings who are born to the same parents in the same year and month. Unfortunately, we cannot differentiate between identical and fraternal twins. We know that opposite sex twins are fraternal, but same sex twins can be either fraternal or identical. We have a large sample of twins – more than 100,000 pairs of twins with longevity data.

Table 8 shows the results. Brother-brother correlations in lifespan of 0.18 (Column 1) are larger among twins than among siblings (0.13 in Table 6) as predicted by the model. Similarly, sister-sister twin correlations (0.16) are larger than siblings (compare to 0.10 in Table 6). So are the correlations in education and household income. This is true even though the intergenerational correlations in longevity with the father are almost the same in both samples. Among twins, we also find that correlations in education and household income are much larger than correlations in lifespan. But these twin correlations in longevity are still below 0.20 (for siblings of the same sex) suggesting that the genetic heritability of lifespan is low, and that random shocks to longevity are large, a point to which we return below. This is similar to the findings using previous twin designs which find that genes explain about 25% of the variation in lifespan (see review by Dato et al. 2017).

c. Comparing correlations in longevity and education

Our results suggest an important role of circumstances in the transmission of longevity. While we don't know which environmental factors explain the correlations we observe, one hypothesis is that the same factors that lead to higher transmission of socio-economic advantage would lead to higher transmission of health. Areas with more equality of resources—which may lower the intergenerational persistence—likely spend time and resources to both educate their children and keep them healthy. If this is the case, we might expect that in places where education or income is transmitted from one generation to the next, health would also be transmitted as well. Intergenerational correlations in education, income and health might also be high in the same locations because education, income, and adult health are highly correlated. Our

hypothesis then is that correlations in SES and correlations in longevity will be positively related.

We use our sibling sample to estimate within sibling correlations in education and longevity by birth state or by cohort to investigate this hypothesis. In Figure 6, we plot the correlation in longevity (y-axis) across states (or across cohorts) against the correlation in education (x-axis). We present the results separately for sisters, brothers, and brother-sister pairs. The plots show that the places where the correlation in education is large are not the same places where the correlation in longevity is high. A simple regression finds that the slopes of most of these relationships are small and statistically insignificant.

Focusing on the correlations across state of birth, for brothers, the sibling correlation coefficient in education ranges from 0.44 to 0.62. The correlation coefficient in longevity has substantially less variation, ranging from around 0.10 to 0.18. The regression of the longevity coefficients on the education coefficients has a slope of -0.011, nearly zero and not statistically significant. A similar regression for sisters yields statistically significant but small positive slope of 0.047, and the slope for male-female correlations is actually negative.

The results are similar when we look at cohort-based comparisons: cohorts with high sibling correlations in education do not generally have high sibling correlations in longevity.[51] Families and communities that succeed in replicating their success in one domain do not necessarily do so in others. This suggests that mobility is an outcome specific process, that depends on inputs that are outcome-specific. This result is consistent with recent work in this area.[52]

d. Interpretation, comparison with previous estimates, and discussion

[51] An exception is that the intergenerational correlations education and longevity are negatively related when we look at brother-sister pairs; we are unsure as to why we observe this relationship.

[52] These findings are consistent with the results in Fletcher and Jajtner (2021) who investigate the same question using a contemporary sample of about 16,000 children from the AddHealth survey and mobility measures for many outcomes. (Their health outcomes include self-reported health status, obesity, smoking, and alcohol drinking.) They conclude that "people and places with high mobility in one domain are not necessarily highly mobile in other domains." Halliday et al. (2021) use the PSID and compare mobility in self-reported health and in incomes. They also find that the two measures are only weakly related, with mobility in health being larger than mobility in income. This suggests that mobility is an outcome specific process, that depends on inputs that are outcome-specific. A natural next step in this research agenda would be to correlate the IGPL by place or cohort with specific measures of inputs into health or education.

Appendix Table A1 reviews many of the studies that have estimated the IGPL and sibling/twin correlations, using a variety of different populations and samples. These studies were first conducted at the end of the 19th century and continue to present day. Although our dataset is substantially larger and is likely to be more representative of the populations of interest than the data used in many previous studies, it is striking how similar the estimates are, particularly when we compare to estimates coming from larger samples and making similar restrictions, as shown in Figure 7.[53]

Our results, combined with these earlier estimates, suggest that the IGPL (measured as a regression coefficient) in longevity is low, in the range of 0.05 to 0.25. While IGPL does indeed capture genetic influences, the findings also suggest that the genetic component of longevity, while significant, is of moderate size for the population as a whole. Recent research looking to identify specific genes that affect longevity has produced estimates that are significantly lower, consistent with the idea that even twin studies overstate the contribution of genes to outcomes. The most recent GWAS study identified only 12 SNPs (out of millions of possible SNPs candidates) that affect longevity (Timmers et al. 2019). They report that an increase of one standard deviation in the polygenic score (weighted average) constructed using these genes increases lifespan by 0.8 to 1.1 years—alternatively they find a 5 years-of-life difference between top and bottom deciles of the polygenic score. While this is a significant effect, it is modest relative to the standard deviation in lifespan (in our data, conditional on surviving to 25, the standard deviation in lifespan is about 16 years).

Why is the IGPL so low? Given the large intergenerational persistence in socio-economic status, and the fact that we know specific genes that carry disease are transmitted within families, one might expect that the IGPL would be larger, and possibly even larger than correlations in SES measures. We consider two explanations for this finding here.

The first is the possibility that longevity is poorly measured and as a result the correlations are weak. This seems an unlikely possibility, as we don't observe the IGPL increasing much since 1880 despite vast improvements in the vital registration system and the increasingly widespread availability of death certificates. To rule out measurement error, we re-

[53] For example, Kaplanis et al. (2018) find a coefficient of 0.12 when regressing child longevity on the average of parental longevity. We find a coefficient of 0.14.

estimate our intergenerational correlations and sibling correlations on a portion of our sibling sample where we include only sibling pairs in which both siblings have a death record attached to their family tree (Appendix Table A8). Despite the smaller sample size, the IGPL is basically unchanged. However, this result may be due to low quality in the father's death certificate information. If we compare sibling correlations instead, we do see a small (roughly 10%) increase in the correlations in the sample with certificates. Thus, the correlations may be somewhat underestimated due to measurement error, but this error is not large enough to produce correlations in the range of the SES correlations we observe.

A second explanation is that the stochastic component of longevity is larger than what it is for other outcomes. Indeed, in a theoretical paper on this issue, Vaupel (1988) demonstrates using a simulation of his frailty (health) model that, even if the correlation in (unobserved) frailty between parent and children is exactly equal to 1, the correlation in longevity can be close to zero.[54] So even if frailty is a large determinant of longevity, if the stochastic component of longevity is large, the IGPL will be low.

To investigate this, we conduct a variance decomposition of education and longevity among siblings. Genetic influences and parental investments that are common among all children are generally captured using family fixed effects or random effects. We can assess how these fixed effects predict longevity and education as a means of placing an upper bound on the influence of the family on these outcomes, as in Björklund and Salvanes (2011).

Table 9 presents the results for our sibling sample. Among all siblings the correlation in lifespan is roughly 0.10, and the correlation in education is 0.55 (panel A). If we regress longevity on observables, not including parental or siblings' longevity, then the R-squared of the regression is low, 0.04 for longevity and 0.13 for education (panel B). If we include family fixed effects (to account for genes and for parental investments) the R-squared increases substantially for education (to 0.73) but much less so for longevity (0.38). If we assumed that the fixed effects capture the effects of genes, then we can put an upper bound on the role of genes as explaining 34% of the variance for longevity and 60% of the variance for education. Of course, the fixed effects capture more than genes, they also capture parental SES/investments and shared environments. This exercise demonstrates that families have a more limited influence on

[54] Frailty is a measure of the distribution of health or disease susceptibility of an individual which in turn determines their probability of dying.

longevity than they have on education. The R-squared in each panel also shows that the observables (including the family effects) predict a substantially smaller portion of the longevity variance, but a large portion of the education variance. Altogether these findings suggest that Vaupel's (1988) hypothesis is consistent with the data: the IGPL is low because the stochastic component of longevity is large, and larger than for SES outcomes.[55]

VIII. Conclusion

While there is a robust literature examining the intergenerational correlation across a broad range of economic outcomes such as income and wealth, much less attention has been paid to the correlation in health, primarily due to an absence of data. In this paper, we use newly available data from family trees that include over 26 million individuals living in the United States from 1900 to 1920 and their parents to estimate the intergenerational correlations in longevity. Consistent with research in other fields, we find small correlations in longevity, with father-son correlations in the range of 0.09, and correlations with both parents on the order of 0.14. Taking advantage of data on the family structure, we use estimates of sibling correlations to show that these low correlations in longevity are in stark contrast to the correlations of other outcomes such as education, occupation, and income using the same sample and methods. Our data suggest that these longevity correlations are low (and lower than SES correlations) because the stochastic component of lifespan is large relative to the contribution of family environments, which is not true for SES measures. Importantly, we find this correlation remains low over time, suggesting limited evidence of aggregate effects of technological and medical advances over this time period. We also document that sibling correlations in lifespan across states or cohorts are very weakly related to correlations in education, suggesting that the factors that determine the IGC in lifespan will be different from those that determine the IGC in education.

Several questions remained unanswered. First, our data does not appropriately represent immigrants, non-whites and short-lived individuals. The fact that correlations in longevity are

[55] These findings also suggest that another interesting avenue for future research is to estimate the correlation in the underlying risk (or frailty) that is transmitted across generations; longevity is clearly a poor proxy. In this area, the work by Halliday et al. (2021) is informative. They correlate overall measures of health across parents and children. They find a correlation in the 0.3 range, larger than ours but still significantly smaller than correlations in education or income. Thus, perhaps correlations in frailty are larger than correlations in longevity, but nevertheless it would appear that health mobility is larger than SES mobility.

low within our study and similar to those found in other studies of different populations suggest that these correlations would also be low for these groups, but this would need to be verified. Second, we uncover some variation in time and space in the persistence of longevity. However, it is not clear what drives this variation. We find large increases in the persistence of longevity in the early 19th century, similar to documented increases in the persistence of income during the same period. But the sources of this increase are unclear. Similarly, we find some variation in space in the persistence of longevity but we do not know why some places are more conducive to persistence than others. Given that persistence in longevity and education are broadly uncorrelated, we can only say that these factors are likely to be specific to longevity, but we haven't identified which factors increase or decrease persistence in longevity. The finding that education, income and longevity can persist, but their persistence is unrelated suggests that efforts to combine them would yield a more complete picture of how wellbeing is transmitted over generations. Our findings suggest that the transmission of wellbeing across generations is larger for more recently born cohorts but, overall, lower than measures of SES would suggest.

References

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.

Amin, V., Böckerman, P., Viinikainen, J., Smart, M. C., Bao, Y., Kumari, M., ... & Pehkonen, J. (2017). Gene-environment interactions between education and body mass: Evidence from the UK and Finland. *Social Science & Medicine*, *195*, 12-16.

Andersen, C. (2021). Intergenerational health mobility: Evidence from Danish registers. *Health Economics*, *30*(12), 3186-3202.

Anderson, D. M., Charles, K. K., Rees, D. I., & Wang, T. (2021). Water purification efforts and the black-white infant mortality gap, 1906–1938. *Journal of Urban Economics*, *122*, 103329.

Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, *115*(42), E9765-E9772.

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

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.

Becker, G. S., Kominers, S. D., Murphy, K. M., & Spenkuch, J. L. (2018). A theory of intergenerational mobility. *Journal of Political Economy*, *126*(1), 7-25.

Beeton, M., & Pearson, K. (1901). On the inheritance of the duration of life, and on the intensity of natural selection in man. *Biometrika*, *1*(1), 50-89.

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., ... & Poulton, R. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological science*, *27*(7), 957-972.

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.

Biavaschi, C., Giulietti, C., & Siddique, Z. (2017). The economic payoff of name Americanization. *Journal of Labor Economics*, *35*(4), 1089-1116.

Björklund, A., & Jäntti, M. (2012). How important is family background for labor-economic outcomes?, *Labour Economics*, *19*(4), 465-474.

Björklund, A., & Jäntti, M. (2020). Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: A comparison of four approaches. *Research in Social Stratification and Mobility*, 100455.

Björklund, A., & Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In *Handbook of the Economics of Education* (Vol. 3, pp. 201-247). Elsevier.

Björkegren, E., Lindahl, M., Palme, M., & Simeonova, E. (2022). Pre-and Post-Birth Components of Intergenerational Persistence in Health and Longevity Lessons from a Large Sample of Adoptees. *Journal of Human Resources*, 57(1), 112-142.

Black, Sandra E. & Devereux, Paul J. (2011). Recent Developments in Intergenerational Mobility, Handbook of Labor Economics, in: O. Ashenfelter & D. Card (ed.), Handbook of Labor Economics, edition 1, volume 4, chapter 16, pages 1487-1541, Elsevier.

Black, S. E., Grönqvist, E., & Öckert, B. (2018). Born to lead? The effect of birth order on noncognitive abilities. *Review of Economics and Statistics*, *100*(2), 274-286.

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

Bleakley, H. (2010). Malaria eradication in the Americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, *2*(2), 1-45.

Blossfeld, H. P. (2009). Educational assortative marriage in comparative perspective. *Annual Review of Sociology*, *35*, 513-53

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

Charles, K. K., & Hurst, E. (2003). The correlation of wealth across generations. *Journal of Political Economy*, *111*(6), 1155-1182.

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.

Classen, T. J. (2010). Measures of the intergenerational transmission of body mass index between mothers and their children in the United States, 1981–2004. *Economics & Human Biology*, *8*(1), 30-43.

Classen, T. J., & Thompson, O. (2016). Genes and the intergenerational transmission of BMI and obesity. *Economics & Human Biology*, *23*, 121-133.

Costa, D. L. (2015). Health and the Economy in the United States from 1750 to the Present. *Journal of Economic Literature*, *53*(3), 503-70.

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

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2011). *Socioeconomic status and health: dimensions and mechanisms*. The Oxford Handbook of Health Economics, Glied, S. and Smith P. (ed.) Oxford University Press 2011: pp. 124-163

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

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

Dato, S., Rose, G., Crocco, P., Monti, D., Garagnani, P., Franceschi, C., & Passarino, G. (2017). The genetics of human longevity: an intricacy of genes, environment, culture and microbiome. *Mechanisms of ageing and development*, *165*, 147-155.

Deelen, J., Evans, D. S., Arking, D. E., Tesi, N., Nygaard, M., Liu, X., ... & Ware, E. B. (2019). A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nature Communications*, *10*(1), 1-14.

Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). "Genetic and educational assortative mating among US adults." *Proceedings of the National Academy of Sciences*, *111*(22), 7996-8000.

Ferrie, J., Massey, C., & Rothbaum, J. (2021). Do grandparents matter? multigenerational mobility in the united states, 1940–2015. *Journal of Labor Economics*, *39*(3), 597-637.

Fletcher, J., & Jajtner, K. M. (2021). Intergenerational health mobility: Magnitudes and Importance of Schools and Place. *Health Economics*, *30*(7), 1648-1667.

Fletcher, J. M., Lu, Q., Mazumder, B., & Song, J. (2023). Understanding Sibling Correlations in Education: Molecular Genetics and Family Background (No. 15862). Institute of Labor Economics (IZA).

Fogel Robert W. Nutrition and the decline in mortality since 1700: Some preliminary findings. In: Engerman SL, Gallman RE, editors. Long-Term Factors in American Economic Growth. Chicago: University of Chicago Press; 1986. pp. 439–556.

Fogel, R. W. (2004). Health, nutrition, and economic growth. *Economic development and cultural change*, *52*(3), 643-658.

Galama, T. J., Lleras-Muney, A., & van Kippersluis, H. (2018). *The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence* (No. w24225). National Bureau of Economic Research.

Gavrilov, L., & Gavrilova, N. (2001). Biodemographic Study of Familial Determinants of Human Longevity. Population: An English Selection, 13(1), 197-221.

Giuntella, Osea, Giulia La Mattina, and Climent Quintana-Domeque. (2022). *Intergenerational Transmission of Health at Birth: Fathers Matter Too!* (No. w30237). National Bureau of Economic Research.

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

Hacker, David J. Decennial Life Tables for the White Population of the United States, 1790–1900. Historical Methods. 2010;43(2):45–79.

Halliday, T. (forthcoming), "Intergenerational Health Mobility" in Handbook of Labor, Human Resources and Population Economics.

Halliday, T. J., Mazumder, B., & Wong, A. (2019). The intergenerational transmission of health in the United States: A latent variables analysis. *Health Economics*.

Halliday, T. J., Mazumder, B., & Wong, A. (2021). Intergenerational mobility in self-reported health status in the US. *Journal of Public Economics*, *193*, 104307.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith and Alina Verashchagina (2007). "The Interitance of Educational Inequality: International Comparison and Fifty-Year Trends." The B.E. Journal of Economic Analysis & Policy. Volume 7, Issue 2, Article 10.

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

Hollingsworth T.H., 1976, "Genealogy and historical demography", Annales de Demographie Historique, p. 167-170

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

Isaacs, J. B., Sawhill, I. V., & Haskins, R. (2008). Getting ahead or losing ground: Economic mobility in America. Brookings Institution.

Johnston, D. W., Schurer, S., & Shields, M. A. (2013). Exploring the intergenerational persistence of mental health: Evidence from three generations. *Journal of Health Economics*, *32*(6), 1077-1089.

Jones, C. I., & Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *American Economic Review*, *106*(9), 2426-57.

Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., ... & Bhatia, G. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*, *360*(6385), 171-175.

Kerber, R. A., Obrien, E., Smith, K. R., & Cawthon, R. M. (2001). Familial Excess Longevity in Utah Genealogies. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences,56(3).

Leive, A., & Ruhm, C. J. (2020). *Has Mortality Risen Disproportionately for the Least Educated?* (No. w27512). National Bureau of Economic Research.

Levine, David & Bhashkar Mazumder. (2007). The Growing Importance of Family: Evidence from Brothers Earnings. *Industrial Relations*, Volume 46.

Lindahl, M., Lundberg, E., Palme, M., & Simeonova, E. (2016). Parental influences on health and longevity: lessons from a large sample of adoptees (No. w21946). National Bureau of Economic Research.

Lindahl, M., Palme, M., Massih, S. S., & Sjögren, A. (2015). Long-term intergenerational persistence of human capital an empirical analysis of four generations. *Journal of Human Resources*, *50*(1), 1-33.

Lleras-Muney, A., Price, J., & Yue, D. (2022). The association between educational attainment and longevity using individual-level data from the 1940 census. *Journal of Health Economics*, 84, 102649.

Lundborg, P., & Stenberg, A. (2010). Nature, nurture and socioeconomic policy—What can we learn from molecular genetics?. *Economics & Human Biology*, *8*(3), 320-330.

Luo, Y., LaPierre, T. A., Hughes, M. E., & Waite, L. J. (2012). Grandparents providing care to grandchildren: A population-based study of continuity and change. *Journal of Family Issues*, *33*(9), 1143-1167.

Mare, R. D. (2011). A multigenerational view of inequality. *Demography*, *48*(1), 1-23.

Marmot, M. (2015). The health gap: the challenge of an unequal world. *The Lancet*, *386*(10011), 2442-2444.

Mazumder, B. (2008). Sibling similarities and economic inequality in the US. *Journal of Population Economics*, *21*(3), 685-701.

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.
Oreffice, S., & Quintana-Domeque, C. (2010). Anthropometry and socioeconomics among couples: Evidence in the United States. *Economics & Human Biology*, *8*(3), 373-384.

Papageorge, N. W., & Thom, K. (2020). Genes, education, and labor market outcomes: evidence from the health and retirement study. *Journal of the European Economic Association*, *18*(3), 1351-1399.

Perls, T. T., Wilmoth, J., Levenson, R., Drinkwater, M., Cohen, M., Bogan, H., ... & Puca, A. (2002). Life-long sustained mortality advantage of siblings of centenarians. *Proceedings of the National Academy of Sciences*, *99*(12), 8442-8447.

Piraino, P., Muller, S., Cilliers, J., & Fourie, J. (2014). The transmission of longevity across generations: The case of the settler Cape Colony. Research in Social Stratification and Mobility, 35, 105-119

Preston, S. H., & Haines, M. R. (1991). The social and medical context of child mortality in the late nineteenth century. In *Fatal Years: Child Mortality in Late Nineteenth-Century America* (pp. 3-48). Princeton University Press.

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

Ruby, J. G., Wright, K. M., Rand, K. A., Kermany, A., Noto, K., Curtis, D., ... & Ball, C. (2018). Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics*, *210*(3), 1109-1124.

Schoenmaker, M., de Craen, A. J., de Meijer, P. H., Beekman, M., Blauw, G. J., Slagboom, P. E., & Westendorp, R. G. (2006). Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *European Journal of Human Genetics*, *14*(1), 79-84.

Solon, G. (1999). Intergenerational mobility in the labor market. In *Handbook of labor economics* (Vol. 3, pp. 1761-1800). Elsevier.

Solon, G. (2014). Theoretical models of inequality transmission across multiple generations. *Research in Social Stratification and Mobility*, *35*, 13-18.

Solon, G., Corcoran, M., Gordon, R., & Laren, D. (1991). A longitudinal analysis of sibling correlations in economic status. *Journal of Human Resources*, 509-534.

Song, X., Massey, C. G., Rolf, K. A., Ferrie, J. P., Rothbaum, J. L., & Xie, Y. (2020). Long-term decline in intergenerational mobility in the United States since the 1850s. *Proceedings of the National Academy of Sciences*, *117*(1), 251-258.

Stuhler, Jan. (2018) "A Review of Intergenerational Mobility and its Drivers," JRC Working Papers JRC112247, Joint Research Centre (Seville site)

Thompson, O. (2014). Genetic mechanisms in the intergenerational transmission of health. *Journal of Health Economics*, *35*, 132-146.

Timmers, P. R., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., ... & Kutalik, Z. (2019). Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife*, *8*, e39856.

Troesken, W. (2004). Water, Race, and Disease. Cambridge, MIT Press.

van den Berg, N., Beekman, M., Smith, K. R., Janssens, A., & Slagboom, P. E. (2017). Historical demography and longevity genetics: back to the future. *Ageing Research Reviews*, *38*, 28-39.

van Raalte, A. A., Sasson, I., & Martikainen, P. (2018). The case for monitoring life-span inequality. *Science*, *362*(6418), 1002-1004.

Vaupel, J. W. (1988). Inherited frailty and longevity. *Demography*, *25*(2), 277-287.

Ward, Z. (2021). *Intergenerational mobility in American history: Accounting for race and measurement error* (No. w29256). National Bureau of Economic Research.

Table 1. Summary Statistics

Variable	Full sample	Sons	Daughters
Lifespan	72.9	70.2	76.1
	(16.08)	(15.43)	(16.25)
Father's Average Lifespan	71.7	71.7	71.7
	(13.55)	(13.53)	(13.58)
Mother's Average Lifespan	72.3	72.3	72.3
	(15.89)	(15.84)	(15.96)
Birth year	1901	1901	1901
	(11.62)	(11.62)	(11.62)
White	0.99	0.99	0.99
Black	0.01	0.01	0.01
Northeast	0.15	0.15	0.15
Midwest	0.41	0.41	0.41
South	0.35	0.35	0.35
West	0.07	0.07	0.07
Immigrant Status	0.01	0.01	0.01
Father's Immigrant Status	0.13	0.13	0.13
Mother's Immigrant Status	0.10	0.10	0.10
Number of Siblings	2.89	2.87	2.91
	(2.36)	(2.35)	(2.37)
Birth Order	2.39	2.39	2.40
	(1.68)	(1.68)	(1.69)
Mother's Age at Child's Birth	29.1	29.1	29.2
	(6.71)	(6.69)	(6.73)
Father's Age at Child's Birth	33.9	33.9	34.0
	(8.02)	(7.99)	(8.05)
Education	9.57	9.45	9.70
	(3.12)	(3.21)	(3.01)
Observations	26,134,161	13,944,386	12,189,775

Table 2. Comparison of Matched Sample and Full Census Sample Mean Values

Variable	Matched Analysis Data	Matched Analysis Data (Weighted)	Full Census Sample
Lifespan	73.0	72.3	
Female	0.47	0.47	0.44
Birth Year	1901	1893	1892
White	0.99	0.99	0.86
Black	0.01	0.01	0.13
Northeast	0.15	0.30	0.30
Midwest	0.41	0.30	0.30
South	0.35	0.30	0.30
West	0.07	0.08	0.08
Immigrant	0.01	0.01	0.19
Father is Immigrant	0.13	0.14	0.38
Mother is Immigrant	0.10	0.11	0.36
Observations	26,134,161	26,134,161	38,947,264

Notes: The estimates in this table compare the mean values of individuals who were age 25 or older in one of the US censuses from 1900-1920 based on whether or not we were able to match the individual to information on their lifespan and the lifespan of both of their parents. The weights in Column 2 are based on birth cohort and census region.

Table 3. IGPL for Varying Child and Parent Pairings and Specifications

	Outcome: Lifespan (Years)				
	(1)	(2)	(3)	(4)	(5)
Model	Lifespan (Years) No Controls	(1) + Parent and Child Birth Year FE	(2) + Parent and Child State of Birth FE	(3) + Race and birth order dummies	# of Obs.
Son/Father	0.089	0.090	0.087	0.087	13,944,386
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	
Son/Mother	0.062	0.062	0.059	0.059	13,944,386
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	
Son/Parents' Average	0.140	0.141	0.137	0.137	13,944,386
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	
Daughter/Father	0.075	0.075	0.072	0.072	12,189,775
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	
Daughter/Mother	0.081	0.074	0.071	0.071	12,189,775
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	
Daughter/Parents' Average	0.150	0.142	0.138	0.138	12,189,775
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	
Father/Mother	0.041	0.043	0.040	0.039	10,251,695
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	

Notes: Each cell separately provides the estimated regression coefficient in lifespan, log lifespan, or percentile lifespan between the two individuals indicated in the row. Errors are clustered by family.

Table 4. IGPL for Varying Child and Parent Pairings and Measures

	Outcome				
Model	Lifespan (Years)	Lifespan (Weighted)	Percentile	Log Lifespan	# of obs.
Son/Father	0.090	0.084	0.090	0.076	13,944,386
	(0.0003)	(0.0004)	(0.0003)	(0.0003)	
Son/Mother	0.062	0.056	0.078	0.048	13,944,386
	(0.0003)	(0.0004)	(0.0003)	(0.0003)	
Son/Parents' Average	0.141	0.130	0.162	0.132	13,944,386
	(0.0004)	(0.0005)	(0.0004)	(0.0004)	
Daughter/Father	0.075	0.071	0.079	0.059	12,189,775
	(0.0003)	(0.0005)	(0.0003)	(0.0004)	
Daughter/Mother	0.074	0.072	0.094	0.056	12,189,775
	(0.0003)	(0.0004)	(0.0003)	(0.0003)	
Daughter/Parents' Average	0.142	0.138	0.166	0.128	12,189,775
	(0.0004)	(0.0006)	(0.0004)	(0.0005)	
Father/Mother	0.041	0.033	0.047	0.038	10,251,695
	(0.0003)	(0.0002)	(0.0003)	(0.0003)	

Notes: Each cell separately provides the estimated regression coefficient in lifespan, log lifespan, or percentile lifespan between the two individuals indicated in the row. The only controls included are birth year fixed effects for child, father and mother. Errors are clustered by family.

Table 5. IGPL of Other Family Members

Variable	Bivariate Regressions		Multiple Regressions			
	Son Lifespan	Daughter Lifespan	Son Lifespan	Daughter Lifespan	Son Lifespan	Daughter Lifespan
Father Lifespan	0.099	0.078	0.093	0.073	0.096	0.075
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother Lifespan	0.067	0.075	0.061	0.071	0.063	0.072
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Paternal Grandfather Lifespan	0.024	0.018	0.017	0.0012		
	(0.002)	(0.002)	(0.002)	(0.002)		
Paternal Grandmother Lifespan	0.016	0.016	0.010	0.012		
	(0.001)	(0.002)	(0.001)	(0.002)		
Maternal Grandfather Lifespan	0.028	0.019	0.022	0.013		
	(0.002)	(0.002)	(0.002)	(0.002)		
Maternal Grandmother Lifespan	0.014	0.015	0.008	0.009		
	(0.001)	(0.002)	(0.001)	(0.002)		
Observations	485,402	425,661	485,402	425,661	485,402	425,661

Notes: The sample in each panel is restricted to children for whom we have all grandparents' lifespans available. In the first panel, each cell is a separate regression. In the second panel, each column is a separate regression. The only controls included are birth year fixed effects for child, father, mother, and paternal and maternal grandparents. Standard errors are clustered at the family level.

Table 6. Sibling IGPL for Outcomes in the 1940 Census Compared to Lifespan

Outcome	(1) Lifespan	(2) Education	(3) Income	(4) HH Income	(5) Lifespan	(6) Education	(7) IGPLF
Brother/Brother	0.134	0.554	0.252	0.346	0.134	0.552	0.084
	(0.001)	(0.001)	(0.017)	(0.005)	(0.001)	(0.001)	(0.001)
	3,664,460	3,664,460	3,664,460	3,664,460	4,126,499	4,126,499	4,680,402
Sister/Sister	0.106	0.603	0.171	0.358	0.105	0.594	0.069
	(0.001)	(0.001)	(0.004)	(0.005)	(0.001)	(0.001)	(0.001)
	2,402,338	2,402,338	2,402,338	2,402,338	3,102,766	3,102,766	3,693,559
Sister/Brother	0.035	0.530	-0.110	0.329	0.035	0.526	0.077
	(0.001)	(0.001)	(0.002)	(0.002)	(0.0004)	(0.001)	(0.0004)
	5,747,644	5,747,644	5,747,644	5,747,644	6,988,569	6,988,569	8,183,995

Notes: Each cell in this table is a separate regression of sibling lifespan (or of the indicated outcome) on sibling lifespan (or of the indicated outcome) including birth cohort fixed effects for each person. Errors are clustered by family. In the first four columns, we only use sibling pairs for which information on all four outcomes is available for both siblings. Since occupation and income are often missing for women in the 1940 census, in the next two columns we include all sibling pairs for whom both education and lifespan are available. The final column includes the IGPL between the children in the previous two columns and their fathers.

Table 7. Accounting for SES in the 1940 Matched Sample

Sample	sample with education		sample with education, income and occupation		sample with education		sample with education, income and occupation	
Parental Lifespan	Father				Mother			
Panel A: Son's lifespan								
Parental Lifespan	0.080	0.079	0.078	0.078	0.055	0.052	0.052	0.052
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Child's Education		0.246	0.248	0.267		0.241	0.246	0.267
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Income/100			-0.001	0.006			-0.004	0.003
			(0.001)	(0.01)			(0.001)	(0.001)
Occupation				-0.020				-0.022
				(0.001)				(0.001)
R^2	0.023	0.026	0.026	0.026	0.021	0.024	0.024	0.024
N	7,055,371		6,604,623		7,055,371		6,604,623	
Panel B: Daughter's lifespan								
Parental Lifespan	0.067	0.064	0.064	0.064	0.064	0.058	0.057	0.057
	(0.0004)	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Child's Education		0.391	0.383	0.382		0.374	0.369	0.369
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Income/100			0.017	0.011			0.010	0.007
			(0.002)	(0.002)			(0.002)	(0.002)
Occupation				0.004				0.002
				(0.001)				(0.001)
R^2	0.008	0.015	0.015	0.015	0.009	0.015	0.015	0.015
N	6,054,117		5,249,738		6,054,117		5,249,738	

Notes: The sample used in this table consists of all individuals in the main sample that have at least one sibling. Each regression uses the full controls from table 3 in addition to the variables included in this table.

Table 8. IGPL Among Twins

	Lifespan	Education	Household income	Lifespan (Father/Child Coefficient)
Panel A: Brother-Brother				
	0.183	0.636	0.386	0.078
	(0.006)	(0.005)	(0.024)	(0.004)
N	31,335	31,335	31,335	62,670
Panel A: Sister-Sister				
	0.162	0.667	0.429	0.070
	(0.007)	(0.005)	(0.012)	(0.004)
N	28,020	28,020	28,020	56,040
Panel A: Sister-Brother				
	0.050	0.546	0.362	0.062
	(0.005)	(0.005)	(0.012)	(0.003)
N	45,628	45,628	45,628	91,256

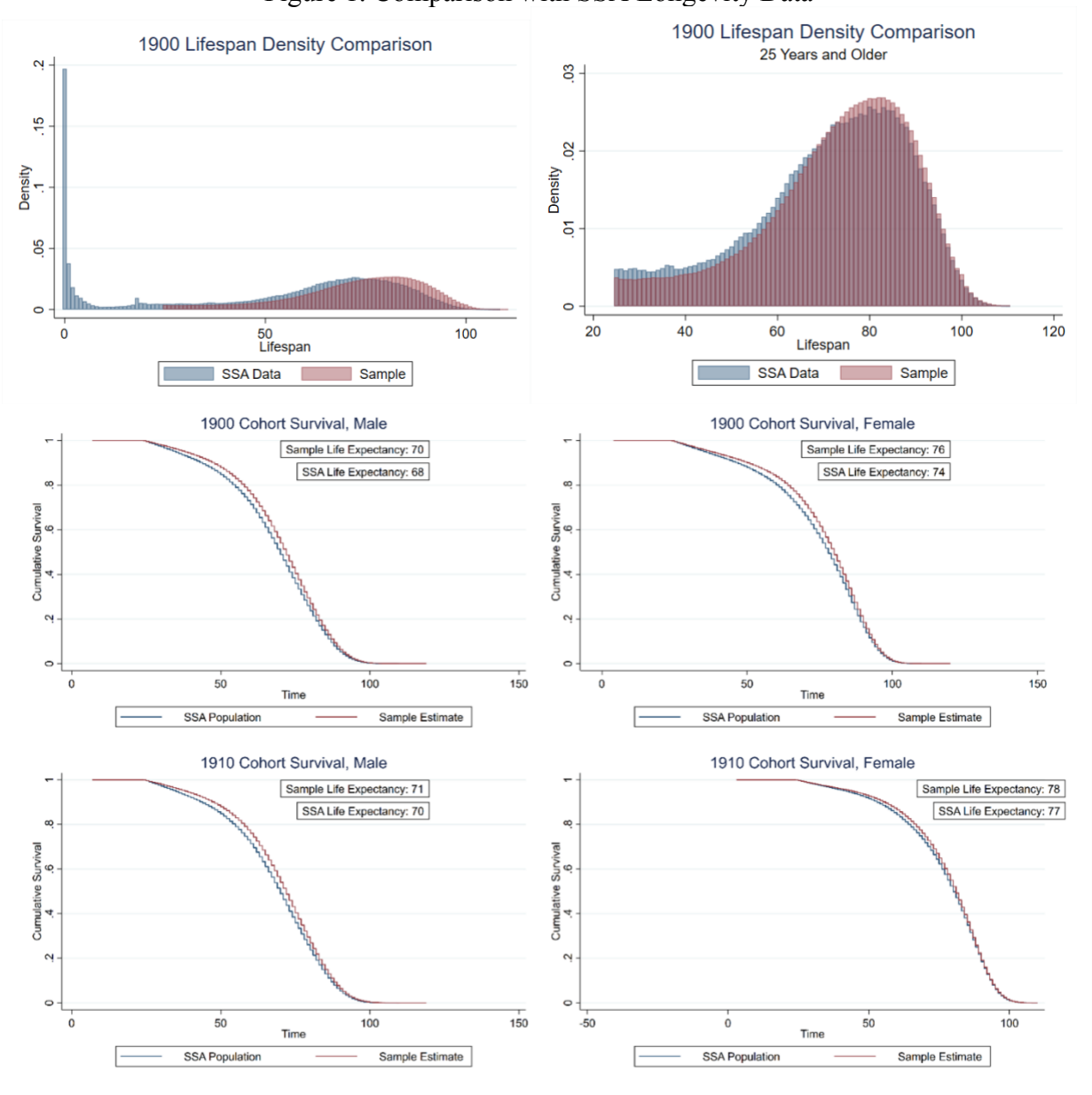
Notes: The sample of twins includes all pairs of individuals born in the same year and month within the same family. Each cell in this table is a separate regression. Each regression includes birth cohort fixed effects for each person. Standard errors are clustered at the family level.

Table 9. Variance Decompositions using Sibling Samples

	Lifespan	Education
Panel A: Raw sibling correlations		
correlation	0.096	0.546
Panel B: Regression without family FE		
R-squared	0.040	0.130
Panel C: Regression with family FE		
R-squared	0.381	0.731
N	22,280,230	13,109,488

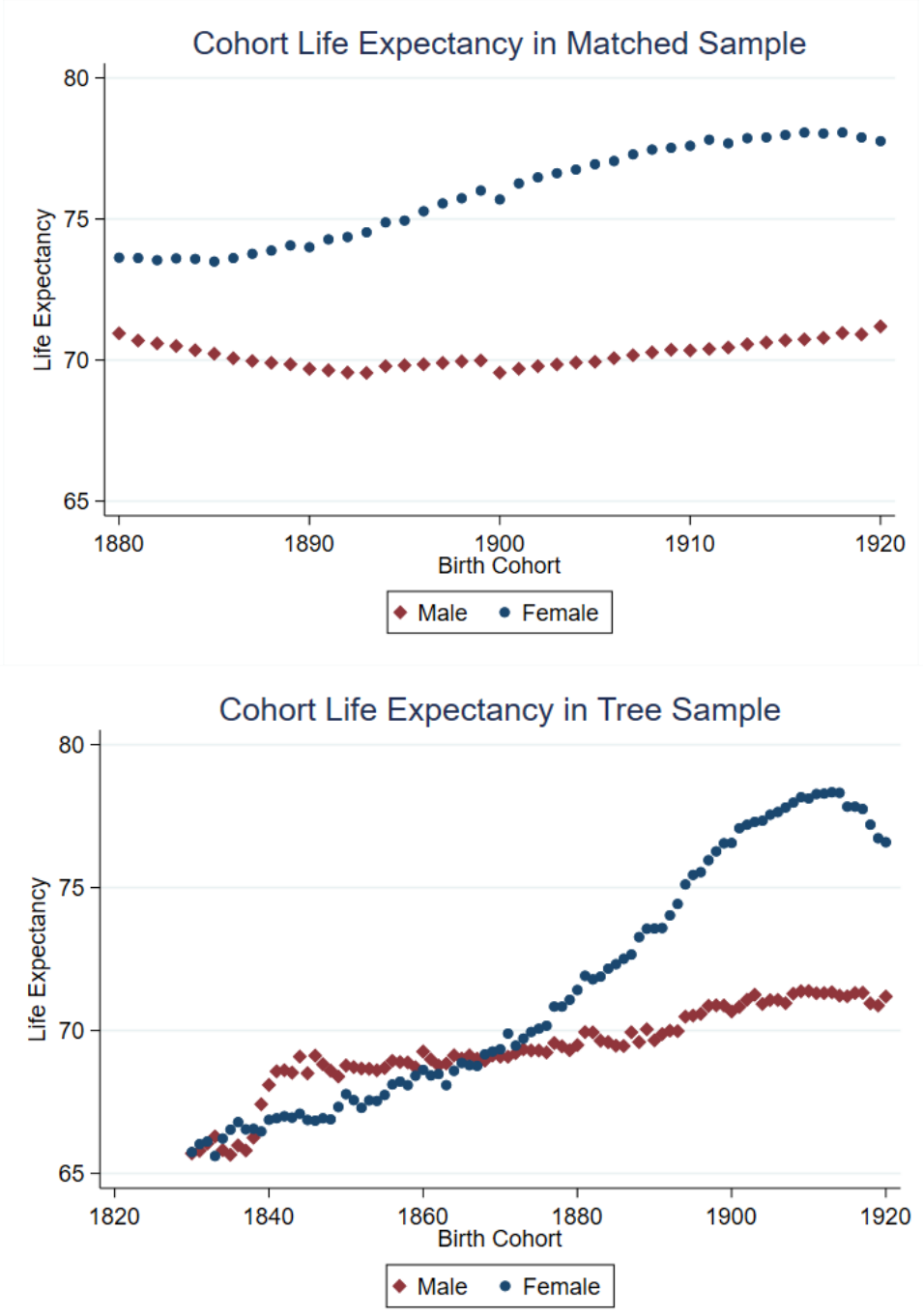
Notes: In this table, we combine all siblings into a single sample. Panel A simply reports that raw sibling correlations in this sample, for reference. Panel B is a regression of the outcome (row header) on covariates: birth cohort of mother FE, birth cohort of father FE, child cohort FE, place of birth FE, indicators for race, gender, number of siblings, birth order, mother and father immigrant status. The regression does not include the siblings' or the parents' longevity. Panel C adds family FE to this regression.

Figure 1. Comparison with SSA Longevity Data



Notes: These figures use our sample derived from the Family Tree (see text for details) and cohort life tables produced by the Social Security Administration (SSA), available here: https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf. Kaplan-Meier estimates are produced using the methods described here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/

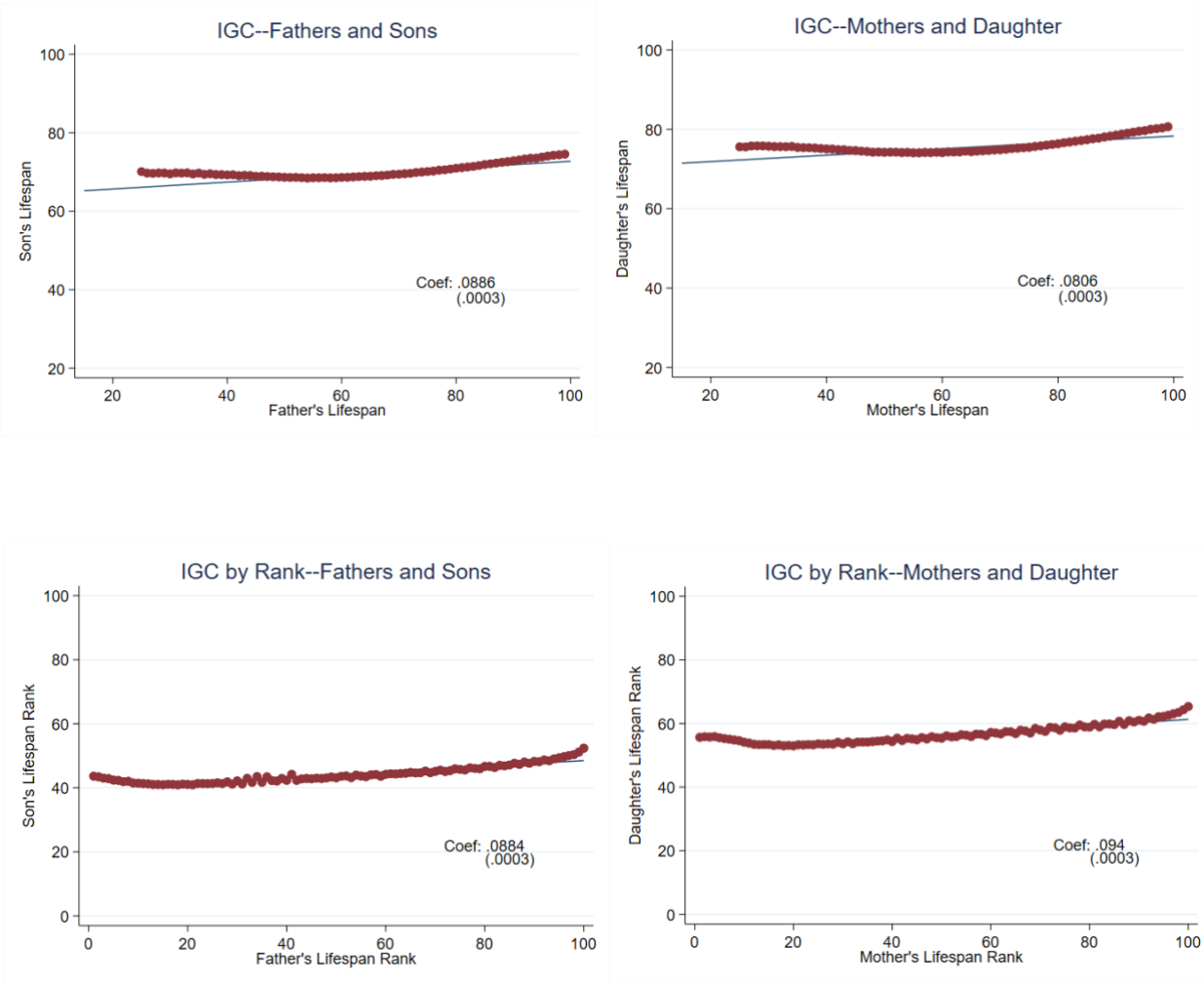
Figure 2. Cohort Life Expectancy (at age 25) Over Time in our Samples



Notes: The top figure shows the evolution of average longevity in our matched sample by cohort, for cohorts born 1880 to 1920 who are observed in the 1900-1920 Censuses. The bottom figure shows mean longevity by cohort, for cohorts from 1830 onwards (there are very few observations before). The data from this figure includes anyone in the family tree with non-

missing data in 2018 when the snapshot of the tree was made. The representativeness of the second data is not well understood.

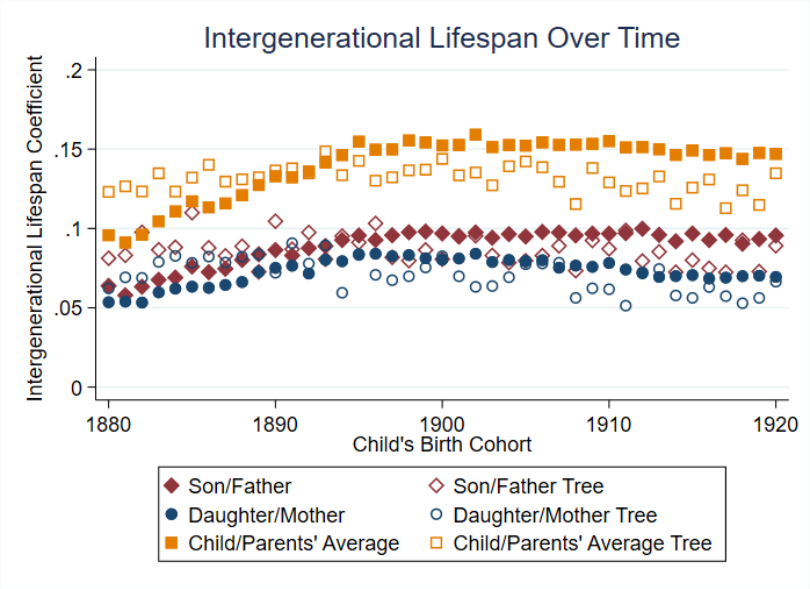
Figure 3. Test for Linearity of Intergenerational Correlations in Lifespan, by Gender



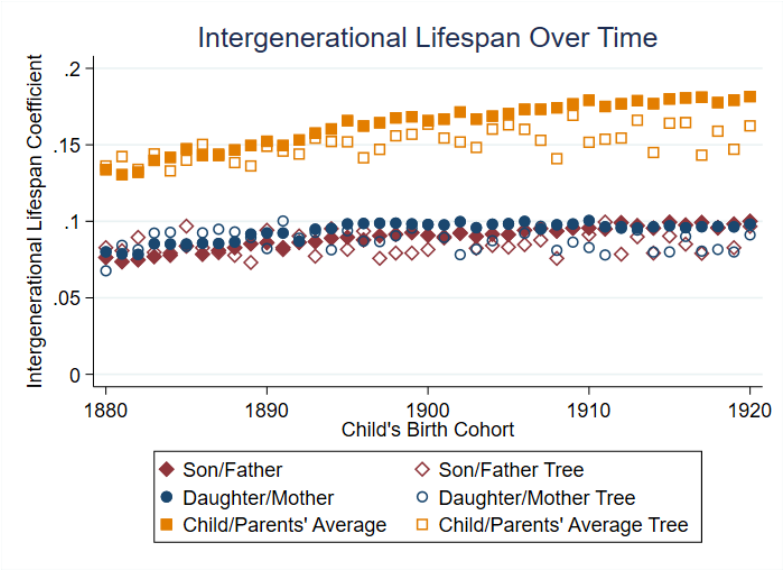
Notes: The top figures provide average of the son's lifespan in one-year bins based on the father's lifespan. All lifespan correlations are conditional on both parents and children living to age 25. The bottom figures relate the average son's (daughter's) percentile in the distribution of the age at death among sons, relative to the father's (mother's) percentile.

Figure 4. Changes in the IGPL Over Time by Level and Rank

a. Level



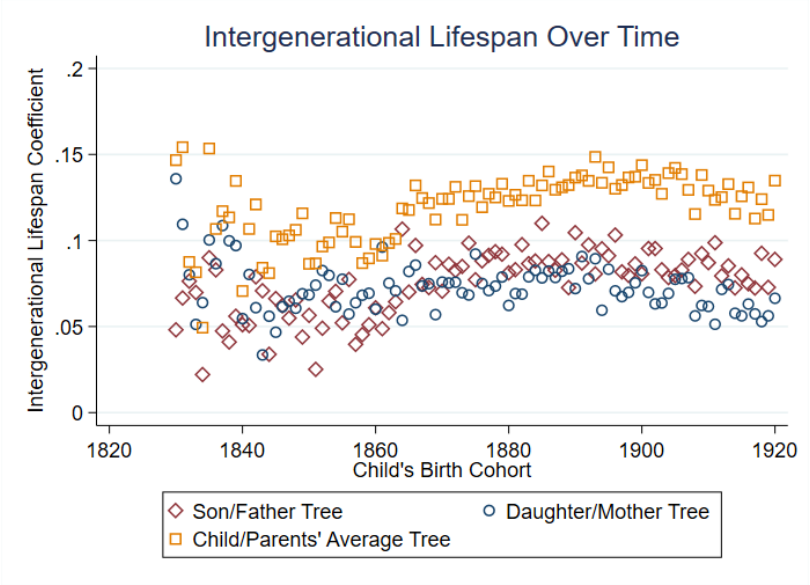
b. Rank



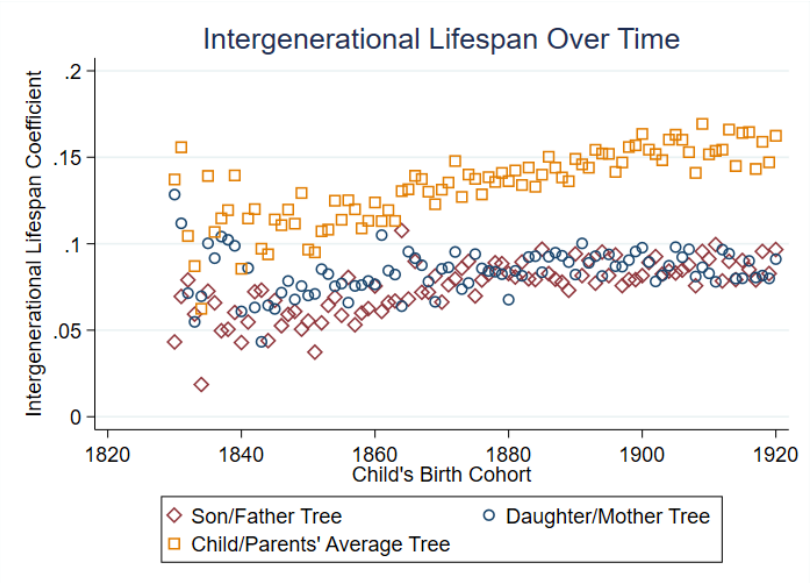
Notes: In this figure, for each birth cohort, we estimate the intergenerational coefficient of lifespan for son/father pairs, daughter/mother pairs, and the correlation between the child's lifespan and the average lifespan of both parents. Each point corresponds to the regression coefficient of a regression of the sons' lifespan on the father's lifespan controlling for birth cohort fixed effects for the parent. We estimate the regression separately for each birth cohort. The solid markers correspond to the results using the data that was matched to the censuses (our primary sample) and the hollow markers correspond to the results we obtain using the entire family tree. Intergenerational lifespan coefficients are estimated using those observations in our dataset in which both parents and children lived to at least age 25.

Figure 5. Long-Term Trends in IGPL by Level and Rank

a. Level



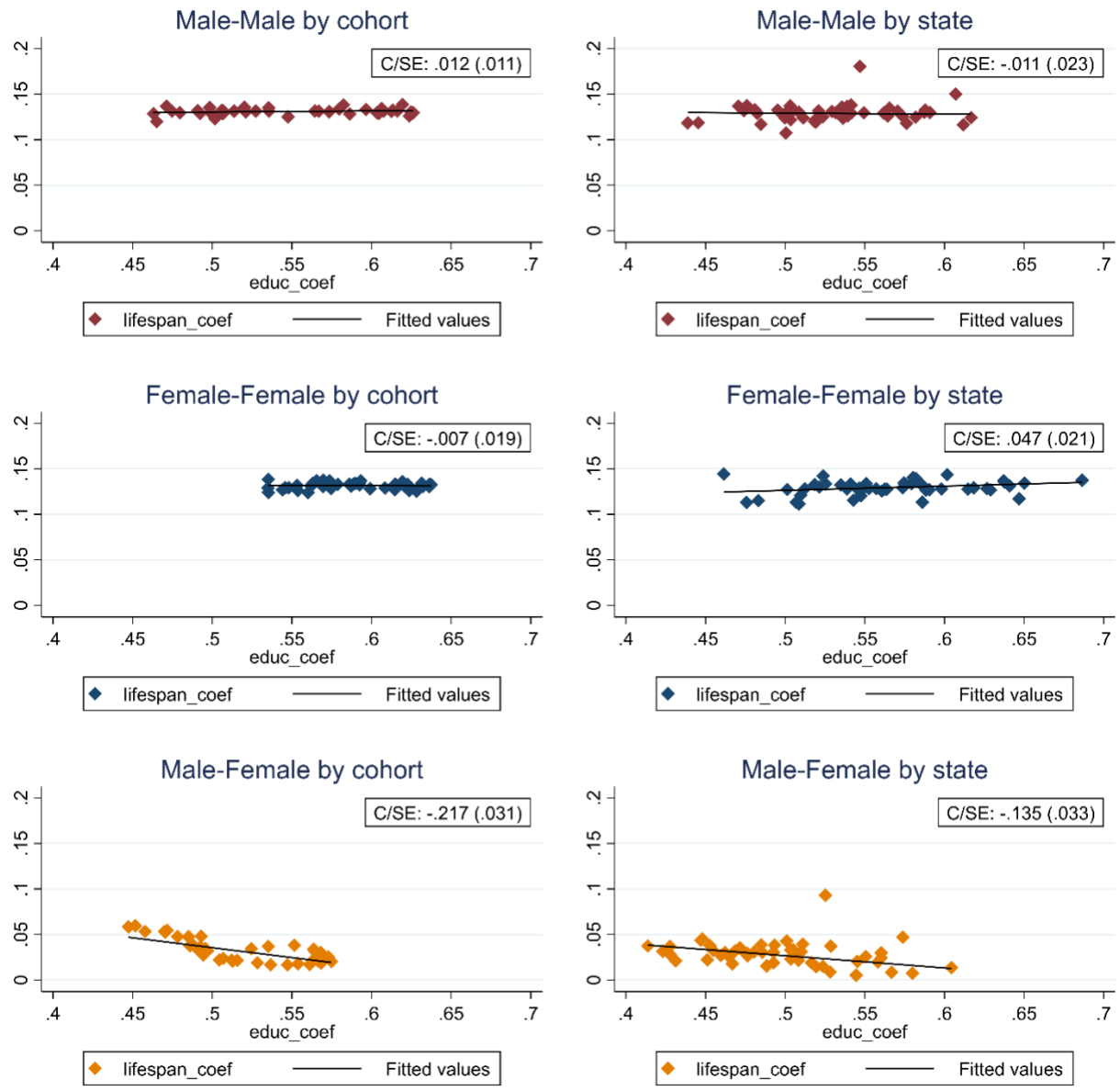
b. Rank



Notes: In this figure, for each birth cohort, we estimate the intergenerational correlation of lifespan for son/father pairs, daughter/mother pairs, and the correlation between the child's lifespan and the average lifespan of both parents. Each regression includes birth cohort fixed effects for child and parent. The solid markers correspond to the results using the data that was matched to the censuses (our primary sample) which only includes cohorts born 1880 to 1920. The hollow markers correspond to the results we obtain using an older version of the Family

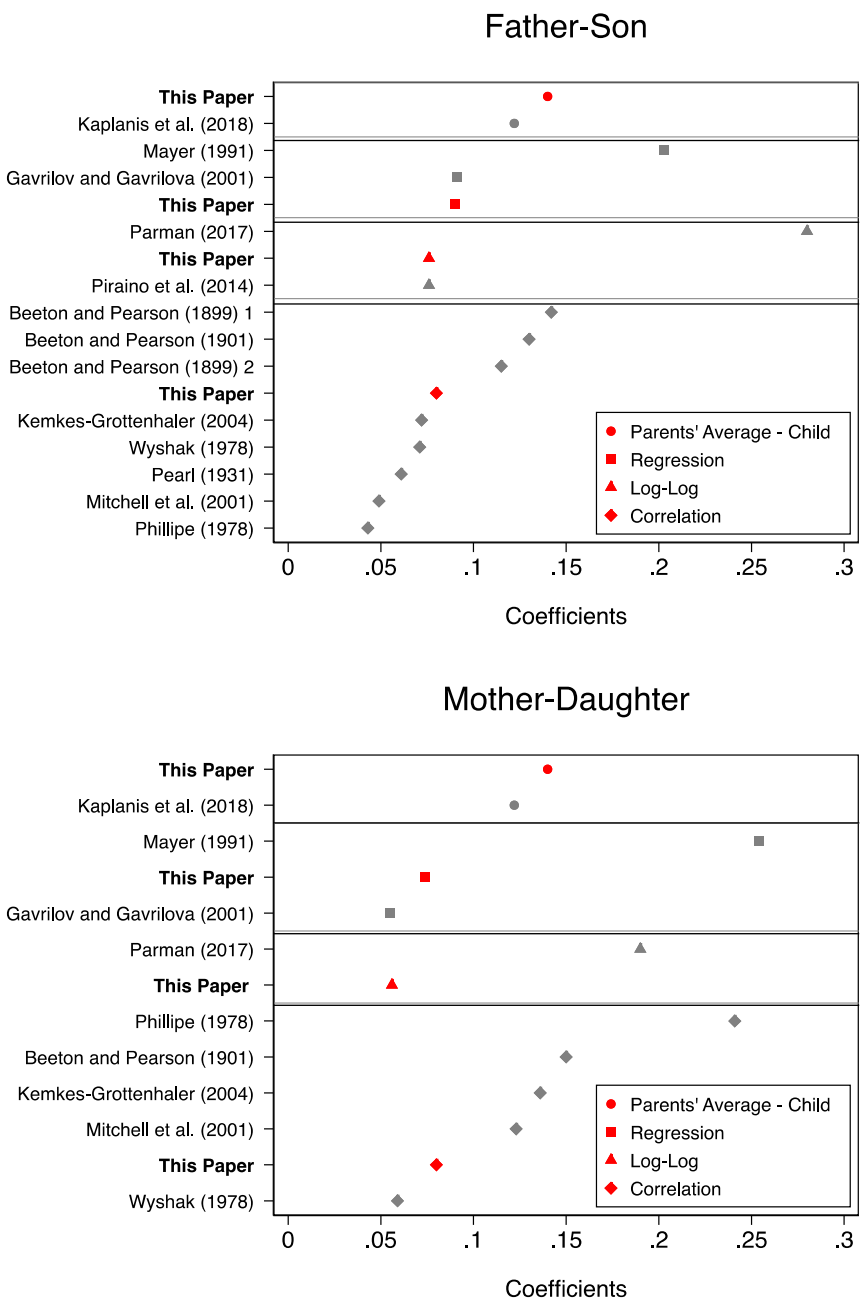
Tree and include all cohorts born since 1830. Lifespan correlations are conditional on parents and children living 25 years.

Figure 6. Sibling Correlations Across Cohorts and States



Notes: These figures plot the sibling correlation in longevity on the y-axis against the sibling correlation in education on the x-axis for a given cohort or state. Fitted lines are weighted by in-sample population of state or cohort respectively. Lifespans are conditional on living to age 25.

Figure 7: Comparison to Previous Estimates of IGPL



Notes: Notes: The figure reports the estimates from various publications. The estimates from this paper come from Table 4 and Appendix Table 3 and refer to correlations/coefficients that do not control for any covariates to make them most comparable to previous estimates. We also report the coefficients that are derived from the age 25+ sample. Estimates from other papers were chosen to be as close as possible as the ones reported here, in terms of the age restrictions and method. Several papers listed in Appendix Table A1 provide estimates that are not directly

comparable and are not included here as a result. Next, we specify the exact location of each estimate in the original publication.

- Father-Son (Mother-Daughter) correlation for Mayer (1991) is calculated in Page 53 (Page 53), which use cohorts of immigrants from England born between 1650 and 1874, without the age restriction.
- Father-Son (Mother-Daughter) correlation for Gavrilov and Gavrilova (2001) is from Table 5 (Table 6), which use 11613 (5025) pairs drawn from European aristocracies born 1800-1880, with the age restriction of surviving until 30.
- Father-Son (Mother-Daughter) correlation for Parman (2017) is from Table 11 (Table 11), which use 585 (424) pairs drawn from cohorts from Meckelenburg County, North Carolina who died between 1934-1975, without the age restriction.
- Father-Son correlation for Piraino et al. (2014) is calculated in Page 112, which use 6059 pairs drawn from cohorts born 1652-1850 in Cape Colony, South Africa, with the age restriction of surviving until 15.
- First estimate for Father-Son correlation for Beeton and Pearson (1899) is calculated in Page 297, which use 1000 pairs drawn from European aristocracies ("Landed Gentry") cohorts, with the age restriction of surviving until 25.
- Second estimate for Father-Son correlation for Beeton and Pearson (1899) is calculated in Page 297,, which use 1000 pairs drawn from European aristocracies ("Peerage") cohorts, with the age restriction of surviving until 20.
- Father-Son (Mother-Daughter) correlation for Beeton and Pearson (1901) is from Table A (Table A), which use 1000 (1064) pairs drawn from cohorts from Britain ("Society of Friends", with the age restriction of surviving until 20.
- Father-Son (Mother-Daughter) correlation for Kemkes-Grottenhaler (2004) is from Table 6 (Table 6), which use 4442 (3885) pairs drawn from cohorts born between 1650 and 1927 in Germany, without the age restriction.
- Father-Son (Mother-Daughter) correlation for Wyshak (1978) is from Table 2 (Table 2), which use 6343 (3125) pairs drawn from cohorts born before 1850 in Salt Lake City, Utah, without the age restriction.
- Father-Son correlation for Pearl (1931) is from Table 11, which use 4407 pairs drawn from cohorts born between 1649 and 1921 in New England, without the age restriction.
- Father-Son (Mother-Daughter) correlation for Mitchell et al. (2001) is from Table 3 (Table 3), which use 709 (586) pairs drawn from cohorts born between 1749 and 1890 in Lancaster County, Pennsylvania, with the age restriction of surviving until 30.
- Father-Son (Mother-Daughter) correlation for Phillipe (1978) is from Table 4 (Table 4), which use 46 (57) pairs drawn from cohorts with parents married between 1820-1899 in Isle-aux-Coudres, Quebec, Canada, with the age at death of offspring before age 20 years.
- The correlation for Kaplanis et al. (2018) is from Supplementary Materials page 13, which use about 130,000 trios of parent-child. It is calculated using parents' average and child longevity and they do not report the correlations for Mother-Daughter and Father-Son pairs. The data come from Geni.com where individual users can upload family tree information.

Appendix Tables and Figures

Table A1: Previous estimates of the intergenerational correlations in lifespan

Paper	IGPL Estimate	SE	Sample size	Population	Cohort
Panel A: Parent Child correlations					
Beeton and Pearson (1899)	Father-Son ("Peerage"): 0.115 Father-Son ("Landed Gentry"): 0.142	Father-Son ("Peerage"): 0.021 Father-Son ("Landed Gentry"): 0.021	Father-Son: 1,000 pairs (Peerage) and 1000 pairs (Landed Gentry)	European aristocracies ("Peerage" and "Landed Gentry")	
Beeton and Pearson (1901)	Father-Son: 0.13 Father-Daughter: 0.13 Mother-Son: 0.13 Mother-Daughter: 0.15	Father-Son:0.02 Father-Daughter:0.02 Mother-Son:0.02 Mother-Daughter:0.02	Father-Son: 1000 pairs Father-Daughter: 1156 pairs Mother-Son: 1220 pairs Mother-Daughter: 1064 pairs	"Society of Friends" from Britain	
Pearl (1931)	Father-Son: 0.061 Father-Daughter: 0.047	Father-Son: 0.01 Father-Daughter: 0.011	Father-Son: 4407 pairs Father-Daughter: 3689 pairs	New England	1649-1921
Wyshak (1978)	Father-Son: 0.071 Father-Daughter: 0.064 Mother-Son: 0.08 Mother-Daughter: 0.059		Father-Son: 6343 pairs Father-Daughter: 3420 pairs Mother-Son: 5505 pairs Mother-Daughter: 3125 pairs	Salt Lake City, Utah	18th and 19th centuries, but born before 1850

Phillipe (1978)	Father-Son: 0.043- 0.129 Father-Daughter: -0.116-0.190 Mother-Son: -0.010-0.194 Mother-Daughter: 0.106-0.241		Father-Son: 128 pairs Father-Daughter: 114 pairs Mother-Son: 134 pairs Mother-Daughter: 132 pairs	Isle-aux-Coudres, Quebec, Canada	parents married 1820-1899
Mayer (1991)	Father-Son: 0.1- 0.3 Father-Daughter: -0.12-0.21 Mother-Son: -0.13-0.32 Mother-Daughter: 0.17-0.21 (shows full 95% CI of estimates)		13,656 individuals	6 New England families who are white, Anglo-Saxon and Protestant immigrants from England	1650-1874
Kerber et al (2001)	Parent-offspring correlation: 0.074		19,575 pairs	Utah	1870-1907
Mitchell et al (2001)	Father-Son: 0.049 Father-Daughter: 0.106 Mother-Son: 0.099 Mother-Daughter: 0.123		Father-Son: 709 pairs Father-Daughter: 610 pairs Mother-Son: 614 pairs Mother-Daughter: 586 pairs	Amish (Lancaster County, Pennsylvania)	1749-1890
Gavrilov and Gavrilova (2001)	Father-Sons: 0.09-0.17 Father-Daughter: 0.06-0.295 Mother-Son: 0.035-0.11 Mother-Daughter: 0.055-0.114	Father-Son:0.01-0.05 Father-Daughter:0.02-0.07 Mother-Son: 0.01-0.05 Mother-Daughter: 0.01-0.07	Father-Son: 11,613 pairs Father-Daughter: 5,025 pairs Mother-Son: 11,613 pairs Mother-Daughter: 5,025 pairs	European aristocracies	1800-1880

Kemkes-Grottenhaler (2004)	Father-Son: 0.051-0.072 Father-Daughter: 0.066-0.13 Mother-Son: 0.059-0.131 Mother-Daughter: 0.103-0.136		Father-Son: 4442 pairs (1015 if 50+) Father-Daughter: 3910 pairs (945 if 50+) Mother-Son: 4404 pairs (1021 if 50+) Mother-Daughter: 3885 pairs (948 if 50+)	Germany	1650-1927
Piraino et al (2014)	Father-Son: 0.173 (0.076 if conditioned on child's survival post 15) Father-Daughter: 0.165 for daughter-father pairs (0.075 if conditioned on child's survival post 15)		Father-Son: 6059 pairs Father-Daughter: 3995 pairs	Cape Colony, South Africa	Born between 1652 - 1850
Parman (2017)	Father-Son: 0.20-0.36 Mother-Daughter: 0.19-0.32	Father-Son: 0.06-0.12 Mother-Daughter: 0.06-0.12	Father-Son: 585 pairs Father-Daughter: 424 pairs	Meckelenburg county, North Carolina	Deaths in 1934-1975 (parents from censuses 1860-1910)
Kaplanis et al (2018)	Parent-child: 0.122	Parent-child: 0.004	Parent-child: 130,000 pairs	US	parents born 1650-1850
Mourits et al (2019)	Offspring of top 10% lived fathers have a survival advantage of 17%, of top 10% of mothers have advantage of 20% and of both parents have 25%		101,577 individuals (16,905 families) Parent-Son: 52367 pairs Parent-Daughter: 49210 pairs	Zeeland province, Netherlands	1812-1886 for children, 1741-1844 for parents

Panel B: Sibling correlations					
Beeton and Pearson (1899)	Brother-Brother: 0.26	Brother-Brother: 0.02	Brother-Brother: 1000 pairs (“Foster’s Peerage” group)	European aristocracies	
Beeton and Pearson (1901)	Brother-Brother: 0.28 Brother-Sister: 0.23 Sister-Sister: 0.33	Brother-Brother:0.02 Brother-Sister: 0.01 Sister-Sister: 0.02	Brother-Brother: 1000 pairs Brother-Sister: 1947 pairs Sister-Sister: 1050 pairs	"Society of Friends"from Britain	
Kerber et al (2001)	Sibling-sibling: 0.107		42,812 pairs	Utah	1870-1907
Phillipe (1978)	Brother-Brother: -0.001-0.263 Brother-Sister: 0.139 Sister-Sister: 0.161-0.315		Brother-Brother: 125 pairs Brother-Sister: 176 pairs Sister-Sister: 110 pairs	Isle-aux-Coudres, Quebec, Canada	parents married 1820-1899
Piraino et al (2014)	Brother-Brother: 0.153 (0.08 if conditioned on survival post 15) Sister-Sister: 0.193 (0.151 if conditioned on survival post 15) Sibling-Sibling: 0.171 (0.086 if conditioned on survival post 15)		122,766	Cape Colony, South Africa	1652 - 1850
Wyshak (1978)	Brother-Brother: 0.077 Sister-Sister: 0.101		Brother-Brother: 5584 pairs Sister-Sister: 2614 pairs	Salt Lake City, Utah	18th and 19th centuries, but born before 1850

Mitchell et al (2001)	Brother-Brother: 0.142 Brother-Sister: 0.082 Sister-Sister: 0.056		Brother-Brother: 700 pairs Brother-Sister: 1416 pairs Sister-Sister: 709 pairs	Amish (Lancaster County, Pennsylvania)	1749-1890
Panel C: Twin correlations					
Herskind et al. (1996)	Male-male twin: 0.26 Female-female twin: 0.23		Male-male MZ twin pairs: 513 Male-male DZ twin pairs: 895 Female-female MZ twin pairs: 520 Female-female DZ twin pairs: 944	Danish same sex twin pairs	1870-1900
Ljunquist et al. (1998)	Male-male MZ twin pairs: 0.33 (reared together), 0.01 (reared apart) Male-male DZ twin pairs: 0.11 (reared together), 0.08 (reared apart) Female-female MZ twin pairs: 0.28 (reared together), 0.15 (reared apart) Female-female DZ twin pairs : 0.12 (reared together), 0.01 (reared apart)	CI: Male-male MZ twin pairs: 0.26-0.39 (reared together), -0.11-0.23 (reared apart) Male-male DZ twin pairs: 0.06-0.15 (reared together), -0.11-0.27 (reared apart) Female-female MZ twin pairs: 0.22-0.34 (reared together), 0.06-0.23 (reared apart) Female-female DZ twin pairs : 0.08-0.15 (reared together), -0.05-0.07 (reared apart)	Male-male MZ twin pairs: 1567 (reared together), 82 (reared apart) Male-male DZ twin pairs: 2814 (reared together), 169 (reared apart) Female-female MZ twin pairs: 1910 (reared together), 97 (reared apart) Female-female DZ twin pairs : 3589 (reared together), 277 (reared apart)	Swedish Twin Pairs	1886-1925

Hjelmborg et al. (2006)	Danish twins: Male-male MZ twin pairs: 0.15 (0.39 if >60) Male-male DZ twin pairs: 0.10 (0.21 if >60) Female-female MZ twin pairs: 0.18 (0.30 if >60) Female-female DZ twin pairs: 0.08 (0.19 if >60) Swedish and Finnish twins: Male-male MZ twin pairs: 0.43 Male-male DZ twin pairs: 0.15 Female-female MZ twin pairs: 0.32 Female-female DZ twin pairs: 0.17	Danish twins: Male-male MZ twin pairs: 0.04 (0.06 if >60) Male-male DZ twin pairs: 0.04 (0.05 if >60) Female-female MZ twin pairs: 0.04 (0.06 if >60) Female-female DZ twin pairs: 0.03 (0.05 if >60) Swedish and Finnish twins: Male-male MZ twin pairs: 0.03 Male-male DZ twin pairs: 0.03 Female-female MZ twin pairs: 0.03 Female-female DZ twin pairs: 0.02	Danish twins: Male-male MZ twin pairs: 851 Male-male DZ twin pairs: 1500 Female-female MZ twin pairs: 862 Female-female DZ twin pairs: 1607 Swedish and Finnish twins: Male-male MZ twin pairs: 829 Male-male DZ twin pairs: 1380 Female-female MZ twin pairs: 987 Female-female DZ twin pairs: 1930	Danish, Finnish and Swedish twins	1870-1910 for Danish births, 1886-1925 for Swedish births, 1880-1910 for Finnish births
Wyshak (1978)	Male on male twin: 0.106 Male on female twin: 0.080 Female on male twin: 0.111 Female on female twin: 0.091		Male on male twin pairs: 2100 Male on female twin pairs: 1224 Female on male twin pairs: 672 Female on female twin pairs: 1059	Salt Lake City, Utah	18th and 19th centuries, but born before 1850

Kerber et al (2001)	Like-sex twins: 0.249 Opposite-sex twins: 0.078		Like-sex twins: 472 pairs Opposite-sex twins:238 pairs	Utah	1870-1907
Panel D: Spousal correlations					
Phillipe (1978)	0.042-0.121		154 pairs	Isle-aux-Coudres, Quebec, Canada	parents married 1820-1899
Parman (2017)	0.142-0.179	0.038-0.047	619 pairs	Meckelenburg county, North Carolina	Deaths in 1934-1975
Mitchell et al (2001)	0.01		312 pairs	Amish (Lancaster County, Pennsylvania)	1749-1890
Wyshak (1978)	0.127		5457 pairs	Salt Lake City, Utah	18th and 19th centuries, but born before 1850
Panel E: Grandparent Correlations					
Kerber et al (2001)	Grandparent-grandchild: 0.015		25,903 pairs	Utah	1870-1907
Piraino et al (2014)	Grandparent-grandchild: -0.022-(-0.012) Great-Grandparent-great-grandchild: 0.021	All insignificant	Grandparent-grandchild: 2601 pairs Great-Grandparent-great-grandchild: 1837 pairs	Cape Colony, South Africa	Born between 1652 - 1850

Table A2. Comparing Tree data with SSA data by cohort

	Sample		SSA		Difference	
Age	Male	Female	Male	Female	Male	Female
1910 Cohort						
25	45.5	51.77	43.34	49.62	2.16	2.15
40	32.84	39.54	31.2	37.64	1.64	1.9
60	17.49	22.92	16.34	21.58	1.15	1.34
80	7.5	9.57	6.86	8.93	0.64	0.64
100	2.05	2.09	1.97	2.25	0.08	-0.16
1900 Cohort						
25	46.64	53.66	45.12	52.07	1.52	1.59
40	33.74	40.58	32.28	39.09	1.46	1.49
60	18.2	23.43	17.12	22.39	1.08	1.04
80	7.27	9.32	7.02	9	0.25	0.32
100	1.36	1.56	1.9	2.19	-0.54	-0.63

Note: Difference calculated (SSA-sample), giving a difference of sample from population. The cohort life tables produced by the Social Security Administration (SSA) are available here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/. Kaplan-Meier estimates are produced using the methods described here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/

Table A3. IGPL for varying child and parent pairings and measures (raw correlations)

	Outcome				
Model	Lifespan (Years)	Lifespan (Weighted)	Percentile	Log Lifespan	Observations
Son/Father	0.08	0.07	0.09	0.06	13,944,386
Son/Mother	0.06	0.06	0.08	0.05	13,944,386
Son/Parents' Average	0.10	0.09	0.12	0.08	13,944,386
Daughter/Father	0.06	0.05	0.08	0.05	12,189,775
Daughter/Mother	0.08	0.06	0.09	0.06	12,189,775
Daughter/Parents' Average	0.10	0.08	0.12	0.08	12,189,775
Father/Mother	0.05	0.04	0.05	0.05	10,251,695

Notes: Each cell separately provides the raw correlation in lifespan, log lifespan, or percentile lifespan between the two individuals indicated in the row.

Table A4. Lifespan quintile transition matrix for mothers and daughters

Daughter Quintile	Mother Quintile 1	2	3	4	5
1	22.42	22.20	20.31	18.55	16.41
2	20.92	21.62	20.68	19.21	17.05
3	20.22	20.57	20.91	20.60	19.29
4	18.96	18.86	19.91	20.92	21.58
5	17.47	16.73	18.18	20.70	25.65

Notes: The sample for this matrix is restricted to mothers and daughters. It compares the portions of the daughter/mother sample in a lifespan quintile given their mother's/daughter's quintile.

Table A5. Lifespan quintile transition matrix for fathers and sons

Son Quintile	Father Quintile 1	2	3	4	5
1	22.88	22.23	20.42	18.70	16.54
2	21.31	21.49	20.61	19.37	17.38
3	19.78	20.07	20.29	20.05	19.18
4	18.67	18.85	19.83	20.73	21.42
5	17.37	17.35	18.85	21.15	25.48

Notes: The sample for this matrix is restricted to fathers and sons. It compares the portions of the son/father sample in a lifespan quintile given their father's/son's quintile.

Table A6. Summary statistics of siblings subsample

	Census based sample matched to FamilySearch, cohorts born 1880-1920	
	Full Sample	Siblings
Average Lifespan	72.97	73.06
	(16.09)	(16.12)
Father's Lifespan	71.66	71.96
	(13.56)	(13.29)
Mother's Lifespan	72.31	72.52
	(15.89)	(15.59)
Birth Year	1901	1901
White	(0.99)	(0.99)
Black	0.01	0.01
<u>Place of birth and ancestry</u>		
Northeast	0.15	0.14
Midwest	0.41	0.41
South	0.35	0.36
West	0.07	0.07
Immigrant Status	0.01	0.01
Father's Immigrant	0.10	0.11
Mother's Immigrant	0.13	0.14
<u>Family characteristics</u>		

Siblings	2.89	3.39
	(2.36)	(2.19)
Birth order	2.39	2.63
	(1.68)	(1.71)
Age mother at birth	33.93	34.13
	(8.02)	(7.89)
Age father at birth	29.13	29.26
	(6.71)	(6.61)
Observations	26,134,160	22,283,088

Notes: The estimates in this table compare individuals who were age 25 or older in one of the US censuses from 1900-1920 for whom we have information about their own lifespan and the lifespan of both of their parents. Standard deviation in parentheses.

Table A7. Sibling correlations for outcomes in the 1940 census compared to lifespan

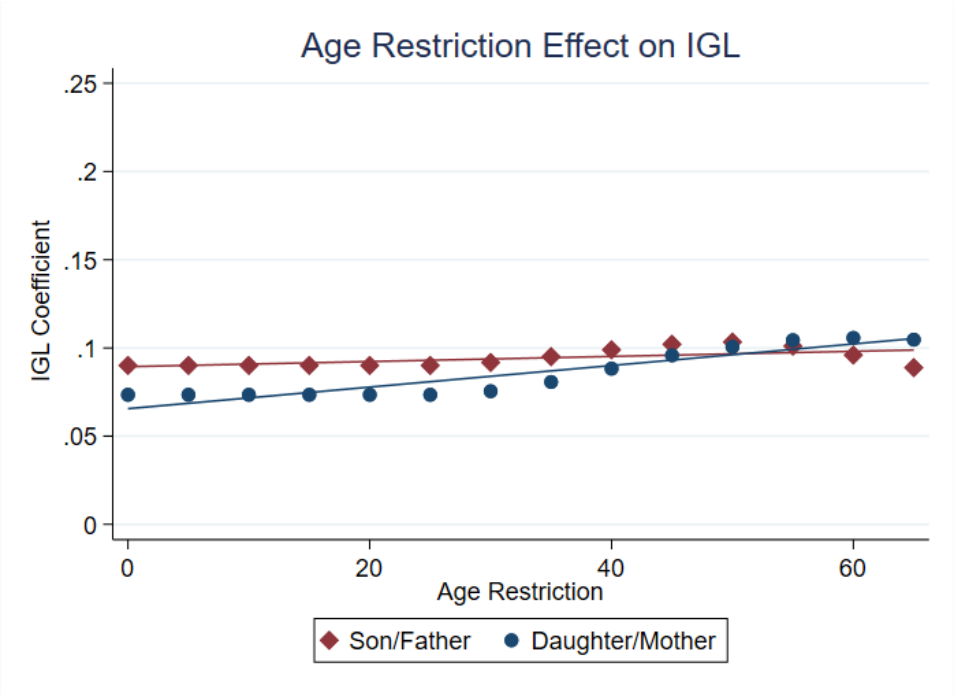
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	Lifespan	Education	Income	HH Income	Lifespan	Education	IGPLF
Brother/Brother	0.141	0.553	0.260	0.347	0.141	0.551	0.094
	3,664,460	3,664,460	3,664,460	3,664,460	4,126,499	4,126,499	4,680,402
Sister/Sister	0.106	0.593	0.167	0.359	0.106	0.585	0.098
	2,402,338	2,402,338	2,402,338	2,402,338	3,102,766	3,102,766	3,693,559
Sister/Brother	0.037	0.529	-0.104	0.328	0.037	0.525	0.094
	5,747,644	5,747,644	5,747,644	5,747,644	6,988,569	6,988,569	8,183,995

Notes: Each cell in this table is a separate correlation. In the first four columns, we only use sibling pairs for which information on all four outcomes is available for both siblings. Since occupation and income are often missing for women in the 1940 census, we include the next two columns and we restrict the sample to just those sibling pairs for which both education and lifespan are available. The final column includes the IGPL between the children in the previous two columns and their fathers.

Table A8: Assessing how the quality of the age at death information affects the results

	All siblings (reproduced from table 6)		All siblings have a death certificate in tree	
	Sibling coefficient	Father coefficient	Sibling coefficient	Father coefficient
Panel A: Sister-sister				
	0.106	0.069	0.118	0.081
	(0.001)	(0.001)	(0.002)	(0.002)
N	2,402,338	3,693,559	229,196	321,367
Panel B: Brother/Brother				
	0.134	0.084	0.159	0.096
	(0.001)	(0.001)	(0.002)	(0.001)
N	3,664,460	4,680,402	542,232	702,565
Panel C: Sister/Brother				
	0.035	0.077	0.060	0.091
	(0.001)	(0.001)	(0.001)	(0.001)
N	5,747,644	8,183,995	659,296	899,945

Figure A1. Age restriction effect on IGPL

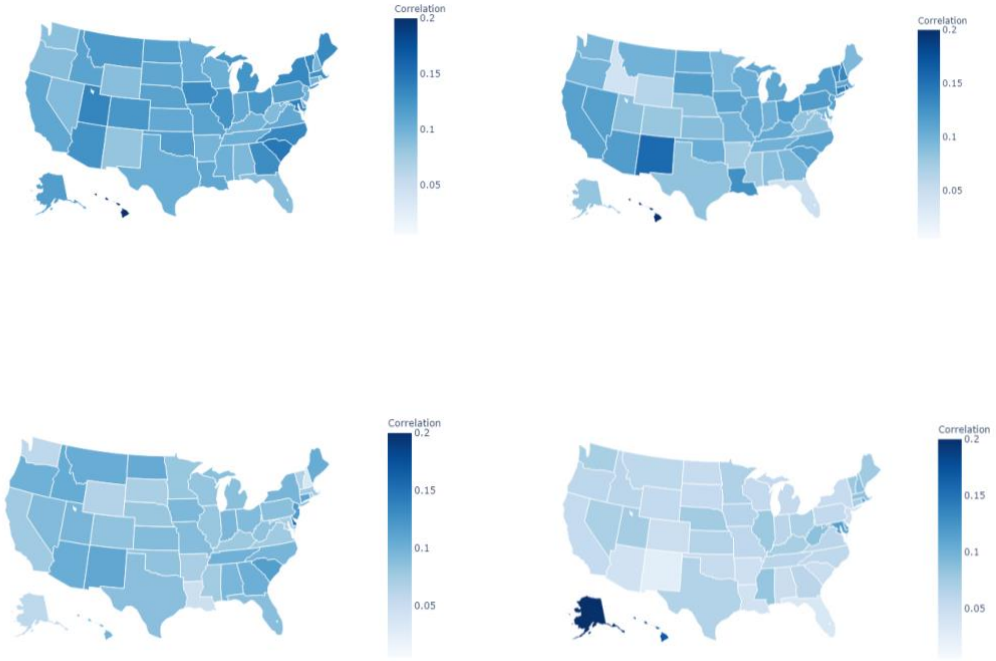


Notes: The specifications in these figures include birth cohort fixed effects for parent, child, and sibling. The age restriction is applied to parent, child, and sibling

Figure A2. Heterogeneity by child's birthplace

1890 Fathers and Sons, 1890 Mothers and Daughters

1920 Fathers and Sons, 1920 Mothers and Daughters



Notes: This figure shows the coefficients from a regression of children's lifespan on parental lifespan without any controls. We estimate these separately by state of birth and birth cohort. The years in the titles are determined by birth year, with 10-year bins beginning in 1880. Coefficients are conditional on parents and children living 25 years.

Appendix A

1 Covariances from Simple Model of Longevity

As outlined in Equation 1 from Section III, the lifespan for an individual i, with gender g, from family j, living in place s, and in cohort c is given by:

$$L^g_{ijsc} = \delta\frac{1}{2}(G^f_{jsc-1} + \eta^f_{ijc} + G^m_{jsc-1} + \eta^m_{ijc}) + \alpha_{sc} + \gamma^g_i + \theta^{SES}_j + e_i$$

Using this model of longevity we can determine the covariance in lifespan between parents, parents and children, siblings, twins, and grandparents and grandchildren.

In this model the **covariance between parents** is given by:

$$\begin{aligned} Cov(L^f_{ijsc-1}, L^m_{ijsc-1}) &= Cov(\delta G^f_{jsc-1} + \alpha_{sc-1} + \gamma^f_i + \theta^{SES}_j + e_i, \delta G^m_{jsc-1} + \alpha_{sc-1} + \gamma^m_i + \theta^{SES}_j + e_i) \\ &= \delta^2 Cov(G^f_{jsc-1}, G^m_{jsc-1}) + V(\alpha_{sc-1}) + V(\theta^{SES}_j) \end{aligned}$$

where the first term denotes the extent to which there is assortative mating between spouses based on genes. The second term is the variance of the (common) environmental component. The third term is the variance in the (common) social economic status component.[1]

The **covariance between father and son** is the given by:

$$\begin{aligned} Cov(L^m_{ijsc-1}, L^m_{ijsc}) = Cov(\ &\delta G^m_{jsc-1} + \alpha_{sc-1} + \gamma^m_i + \theta^{SES}_j + e_i, \\ &\delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{ijc} + \alpha_{sc} + \gamma^m_i + \theta^{SES}_j + e_i) \end{aligned}$$

$$Cov(L^m_{ijsc-1}, L^m_{ijsc}) = \delta^2\frac{1}{2}Cov(G^m_{jsc-1}, G^f_{jsc-1}) + \delta^2\frac{1}{2}V(G^m_{ijsc-1}) + Cov(\alpha_{sc-1}, \alpha_{sc}) + V(\gamma^m_i) + V(\theta^{SES}_j)$$

The covariance between father and son depends on the genetic assortativeness of the parents,

[1] Note that the strict commonality of spouses environment and social economic status could be relaxed and the last two terms could then represent the amount of assortative mating between spouses based on environment and social economic status. While potentially important in many contexts, this simplification does not change the implications discussed in Section III.

the extent to which father and son share genes, the extent to which they share an environment, the variance of the gender component, and the variance in the social economic status component.

Similarly, the **covariance between mother and daughter** is given by:

$$\begin{aligned} Cov(L^f_{ijsc-1}, L^f_{ijsc}) = Cov(\ & \delta G^f_{jsc-1} + \alpha_{sc-1} + \gamma^f_i + \theta^{SES}_j + e_i, \\ & \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{ijc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_i) \end{aligned}$$

$$Cov(L^m_{ijsc-1}, L^m_{ijsc}) = \delta^2\frac{1}{2}Cov(G^f_{jsc-1}, G^m_{jsc-1}) + \delta^2\frac{1}{2}V(G^f_{jsc-1}) + Cov(\alpha_{sc-1}, \alpha_{sc}) + V(\gamma^f_i) + V(\theta^{SES}_j)$$

The **covariance between father and daughter** is the given by:

$$\begin{aligned} Cov(L^m_{ijsc-1}, L^f_{ijsc}) = Cov(\ & \delta G^m_{jsc-1} + \alpha_{sc-1} + \gamma^m_i + \theta^{SES}_j + e_i, \\ & \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{ijc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_i) \end{aligned}$$

$$Cov(L^m_{ijsc-1}, L^f_{ijsc}) = \delta^2\frac{1}{2}Cov(G^m_{jsc-1}, G^f_{jsc-1}) + \delta^2\frac{1}{2}V(G^m_{jsc-1}) + Cov(\alpha_{sc-1}, \alpha_{sc}) + V(\theta^{SES}_j)$$

The covariance between father and daughter is the same as the covariance between father and son except there is no term for the shared gender component. Similarly, the covariance between mother and son is the same as the covariance between mother and daughter except there is no term for the shared gender component.

In addition to the covariances between parents and children we can also express the covariances between different types of siblings. With out loss of generality, let $i = 1$ for one sibling and let $i = 2$ for the another sibling.

The **covariance between male siblings** is given by:

$$\begin{aligned} Cov(L^m_{1jsc}, L^m_{2jsc}) = Cov(\ & \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{1jc} + \alpha_{sc} + \gamma^m_i + \theta^{SES}_j + e_1, \\ & \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{2jc} + \alpha_{sc} + \gamma^m_i + \theta^{SES}_j + e_2) \end{aligned}$$

$$\begin{aligned} Cov(L^m_{1jsc}, L^m_{2jsc}) = & \delta^2\frac{1}{4}V(G^f_{jsc-1}) + \delta^2\frac{1}{4}V(G^m_{jsc-1}) + \delta^2\frac{1}{2}Cov(G^f_{jsc-1}, G^m_{jsc-1}) \\ & + V(\alpha_{sc}) + V(\gamma^m_i) + V(\theta^{SES}_j) \end{aligned}$$

The **covariance between female siblings** is given by:

$$Cov(L^f_{1jsc}, L^f_{2jsc}) = Cov(\, \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{1jc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_1,$$
$$\delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{2jc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_2)$$

$$Cov(L^f_{1jsc}, L^f_{2jsc}) = \delta^2\frac{1}{4}V(G^f_{jsc-1}) + \delta^2\frac{1}{4}V(G^m_{jsc-1}) + \delta^2\frac{1}{2}Cov(G^f_{jsc-1}, G^m_{jsc-1}) + V(\alpha_{sc})$$
$$+ V(\gamma^f_i) + V(\theta^{SES}_j)$$

The **covariance between opposite gender siblings** is given by:

$$Cov(L^m_{1jsc}, L^f_{2jsc}) = Cov(\, \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{1jc} + \alpha_{sc} + \gamma^m_i + \theta^{SES}_j + e_1,$$
$$\delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{2jc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_2)$$

$$Cov(L^m_{1jsc}, L^f_{2jsc}) = \delta^2\frac{1}{4}V(G^f_{jsc-1}) + \delta^2\frac{1}{4}V(G^m_{jsc-1}) + \delta^2\frac{1}{2}Cov(G^f_{jsc-1}, G^m_{jsc-1}) + V(\alpha_{sc}) + V(\theta^{SES}_j)$$

Note that the only difference between male siblings, female siblings, and opposite gender siblings is whether $V(\gamma^m_i)$, $V(\gamma^f_i)$, or no variance in the gender component is included, respectively.

The **covariance between male identical twins** is given by:

$$Cov(L^m_{1jsc}, L^m_{2jsc}) = Cov(\, \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{1jc} + \alpha_{sc} + \gamma^m_i + \theta^{SES}_j + e_1,$$
$$\delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{2jc} + \alpha_{sc} + \gamma^m_i + \theta^{SES}_j + e_2)$$

$$Cov(L^m_{1jsc}, L^m_{2jsc}) = \delta^2\frac{1}{4}V(G^f_{jsc-1}) + \delta^2\frac{1}{4}V(G^m_{jsc-1}) + \delta^2\frac{1}{2}Cov(G^f_{jsc-1}, G^m_{jsc-1})$$
$$+ V(\alpha_{sc}) + V(\gamma^m_i) + V(\theta^{SES}_j) + V(\eta_{ijc})$$

The **covariance between female identical twins** is given by:

$$Cov(L^f_{1jsc}, L^f_{2jsc}) = Cov(\, \delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{1jc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_1,$$
$$\delta\frac{1}{2}(G^f_{jsc-1} + G^m_{jsc-1}) + \eta_{2jc} + \alpha_{sc} + \gamma^f_i + \theta^{SES}_j + e_2)$$

$$Cov(L^f_{1jsc}, L^f_{2jsc}) = \delta^2\frac{1}{4}V(G^f_{jsc-1}) + \delta^2\frac{1}{4}V(G^m_{jsc-1}) + \delta^2\frac{1}{2}Cov(G^f_{jsc-1}, G^m_{jsc-1})$$
$$+ V(\alpha_{sc}) + V(\gamma^f_i) + V(\theta^{SES}_j) + V(\eta_{ijc})$$

Note that the covariances for twins are the same as those for the same gender siblings except that

they also have a term for the variance in the additional shared genetic component.

The **covariance between grandfather and grandson** is given by:

$$\begin{aligned} Cov(L^m_{ijsc-1}, L^m_{ijsc+1}) = Cov(\ & \delta G^m_{jsc-1} + \alpha_{sc-1} + \gamma^m_i + \theta^{SES}_j + e_i, \\ & \delta\frac{1}{4}(G^{ff}_{jsc-1} + \eta^{ff}_{ijc} + G^{mf}_{jsc-1} + \eta^{mf}_{ijc} + G^{fm}_{jsc-1} + \eta^{fm}_{ijc} + G^{mm}_{jsc-1} + \eta^{mm}_{ijc}) \\ & + \delta\frac{1}{2}(\eta^f_{ijc+1} + \eta^m_{ijc+1}) + \alpha_{sc+1} + \gamma^m_i + \theta^{SES}_j + e_i) \end{aligned}$$

$$\begin{aligned} Cov(L^m_{ijsc-1}, L^m_{ijsc+1}) = & \delta^2\frac{1}{4}V(G^{mm}_{jsc-1}) + \delta^2\frac{1}{4}Cov(G^{mm}_{jsc-1}, G^{mf}_{jsc-1}) + \delta^2\frac{1}{4}Cov(G^{mm}_{jsc-1}, G^{ff}_{jsc-1}) \\ & + \delta^2\frac{1}{4}Cov(G^{mm}_{jsc-1}, G^{fm}_{jsc-1}) + Cov(\alpha_{sc-1}, \alpha_{sc+1}) + V(\gamma^m_i) + V(\theta^{SES}_j) \end{aligned}$$

Note, the superscripts mm, mf, fm, and ff represent father's father (i.e. the male male genealogical line), father's mother, mother's father, and mother's mother, respectively. Each of covariances for the other three grandparent-grandchild combinations are analogous to this results, however, with the variance of the gender component ($V(\gamma^m_i)$ or $V(\gamma^f_i)$) omitted for the cross-gender covariances.