NBER WORKING PAPER SERIES

INTERGENERATIONAL TRANSMISSION OF LIFESPAN IN THE US

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, Revised May 2024

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. We are very grateful to Hiram Beltran, Paul Devereux, Jason Fletcher, Tim Halliday, Juliana Londoño-Vélez, Bhash Mazumder, Tom Vogl, and to the seminar participants at the NBER Economic Mobility conference, Chicago mobility conference, the University of Maryland, the University of Melbourne, CUNY, and the University of Wisconsin for helpful comments. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Sandra E. Black, Neil Duzett, Adriana Lleras-Muney, Nolan G. Pope, and Joseph Price. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Intergenerational Transmission of Lifespan in the US
Sandra E. Black, Neil Duzett, Adriana Lleras-Muney, Nolan G. Pope, and Joseph Price
NBER Working Paper No. 31034
March 2023, Revised May 2024
JEL No. I1,I30,J6

ABSTRACT

We examine the transmission of lifespan across generations using a unique dataset containing more than 26 million individuals born between 1880 and 1920. We document new facts about the absolute and relative mobility in lifespan. On average across cohorts, 47% of men and 57 percent of women lived longer than their parents, though this varied across cohorts. Relative measures show higher mobility, with substantially less variation across time and subpopulations. The intergenerational correlation in lifespan (a measure of persistence rather than mobility) is about 0.09 for both sexes – this low correlation is observed across races, education groups, cohorts, and birth states. Finally, we document that the intergenerational persistence of lifespan is much smaller than the persistence in socio-economic status. Moreover, correlations in lifespan and in education are largely independent of each other, suggesting that mobility in well-being may be larger than measures of income alone suggest.

Sandra E. Black
Department of Economics and
School of International and Public Affairs
Columbia University
1022 International Affairs Building
420 West 118th Street
New York, NY 10027
and IZA and also NBER
sblack@columbia.edu

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

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

Nolan G. Pope
University of Maryland
3114 Tydings Hall
7343 Preinkert Dr
College Park, MD 20742
and NBER
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 substantial persistence in economic well-being across generations based on outcomes such as education, income, and wealth.[1] However, much less is known about the persistence of health and longevity across generations, despite the fact that they are also essential determinants of well-being.[2] While we know that there is considerable and rising cross-sectional inequality in health—with richer, more educated individuals being healthier and living substantially longer lives in the US and most other developed countries—it is less well-known how much these differences in health outcomes are transmitted from parents to children.[3] Given that, ultimately, it is the intergenerational correlation in well-being that we care about, understanding the intergenerational persistence in health in addition to the economic components of well-being—as well as how much these components move together—is fundamental.

We use a newly compiled dataset to study the transmission of lifespan—a summary measure of health—across generations. We create our unique dataset by combining United States census records with data from the wiki-style Family Tree created by FamilySearch, which includes over 1.4 billion people—the largest collection of its kind. We start with a dataset of all individuals observed in the 1900, 1910, and 1920 US censuses who were born between 1880 and 1920; we then match these individuals to the FamilySearch database to gather information about their lifespan and the lifespan of their parents, focusing on individuals who survive to at least age 25. Our final sample includes 26.1 million individuals matched to their parents and other family members. This is the largest sample ever constructed to study the intergenerational transmission of health and includes more than 30% of the US population born between 1880 and 1920 that survived to age 25.

These data allow us to document a new series of facts about lifespan transmission in the US. We examine both absolute and relative mobility to provide a comprehensive picture of intergenerational mobility in lifespan over time and across different sub-populations. During the

[1] See Black and Devereux (2011), Stuhler (2018) for summaries of this research.
[2] See Jones and Klenow (2016), and Becker, Philipson, and Soares (2005).
[3] See Cutler et al. (2011), Marmot (2015), Chetty et al. (2016), and Galama et al. (2018) for evidence on the relationship between socioeconomic status and health. Research has shown that, for example, the health gaps between the more and the less educated have grown 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). Other research shows that the variance of lifespan is rising (van Raalte et al. 2018).

late 19th century and the first half of the 20th century, incomes, education and life expectancy were rising rapidly. Yet this was also a period of rising inequality and great economic and political instability. Previous work has shown that, despite these tremendous changes, economic mobility across generations for these cohorts was low and remained stable (Song et al. 2020, Buckles et al. 2023, Ward 2023). We evaluate whether health mobility was similarly low and followed the same trends.

Then, to better understand how lifespan mobility and economic mobility are related, we construct a subsample of siblings alive in 1940 for whom we can observe education, income, and occupation in the 1940 full census in addition to lifespan. (As education and income are not available in previous censuses, it is not possible to estimate intergenerational correlations for these measures for our cohorts.) Using this sibling subsample, we estimate sibling correlations in lifespan, education, and income on a constant sample. These sibling correlations provide an upper bound on intergenerational persistence across these different outcomes.

Finally, we investigate whether these sibling correlations—in lifespan and socioeconomic status—are related across both time and space. Is it the case that areas or periods with high persistence in education also have high persistence in lifespan? One might speculate that in places where socioeconomic status is more persistent from one generation to the next, health would also be more persistent, as factors such as investment in public infrastructure that influence one might also influence the other.

We find that mobility in lifespan is high, no matter how you measure it, although different measures—particularly when comparing absolute versus relative mobility—highlight different patterns across sex and over time, with relative mobility exhibiting much less variation than absolute mobility measures. The intergenerational correlation in lifespan is low and has been low since the 1880s. Moreover, persistence estimates are similarly low across different subgroups, including sex, race/ethnicity, immigration status, place of birth, education, and family types.

When we use a constant sample and look at sibling correlations, we see that correlations in education (~0.5) are higher than correlations in income (~0.3), and correlations in income are higher than correlations in lifespan (~0.15). These results confirm that lifespan and income/SES measures paint a fairly different picture of mobility, even within the same sample. Strikingly, we find that correlations in education and correlations in lifespan are independent of each other in

time and space. 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 lifespan.

Our results suggest that measures of the intergenerational correlation in well-being that incorporate both education/SES and health may show much more mobility than measures of education/SES alone. Additionally, the results also suggest that it is possible to improve well-being mobility without influencing economic mobility. Altogether, economic measures of mobility appear insufficient to characterize well-being mobility across generations.

Finally, we investigate why the intergenerational persistence in lifespan is so low. These small intergenerational estimates were quite surprising, given that it is well-known that there are important genetic determinants of health, and genes are transmitted from parents to children. To do so, we examine twins, where one might expect that the genetic component would be very large. We conclude that, even among twins, the correlations in lifespan are low. In the end, we show that the low correlation in lifespan can be explained by the fact that there is a large stochastic component in the determination of lifespan that is unrelated to family influences such as genetics or shared environments, as hypothesized by Vaupel (1988). This is in stark contrast to economic outcomes such as education or income, which are much more influenced by family environments.

Our paper contributes to a number of literatures in economics and demography on the intergenerational persistence in health. There is a burgeoning literature in economics examining the intergenerational persistence in health, where health is often measured using specific measures of health, such as birthweight, mental health, BMI, or chronic conditions.[4] An obvious limitation of this approach is that these outcomes are only partial measures of health and capture health only at a given age/moment. More recent efforts have focused on constructing health indices that can be compared across generations and that summarize all health aspects at a meaningful point in the lifetime (or across the lifetime) – see the recent handbook chapter by

[4] 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 mothers and children in India is 0.26. Lu and Vogl (2023) find that women whose mothers lost a child are more likely to lose a child.

Halliday (2023) for an excellent summary of this work.[5] Our work focuses on lifespan, a complementary measure of health that focuses on the quantity, and not quality, of life, instead of single measures of health or constructed health indices. Unlike other health and economic measures, lifespan is an objective indicator that is easily compared across individuals and over time, and it does not require age adjustments to account for lifecycle issues.

We also build on this literature by looking at absolute mobility in addition to relative mobility – to our knowledge, measures of absolute mobility in lifespan have not been studied before. In the intergenerational income literature, researchers have shown that measures of absolute and relative mobility can tell very different stories (Deutscher and Mazumder 2023). We show this is true for lifespan as well.

There is a separate large literature in demography investigating the intergenerational persistence in lifespan, which began in the 19th century (e.g. Beeton and Pearson, 1899). Because this literature focuses primarily on identifying genetic influences, it typically uses small convenience samples, making the results hard to generalize, and comparisons across cohorts and heterogeneity analyses difficult.[6] Nevertheless our results are similar, a surprising finding, which is explained by the fact that estimates of persistence are low across all subgroups and time periods we examine.

Finally, we build on the nascent literature examining the relationship between persistence in SES and in health. Using very large samples and measures of lifespan, we confirm what

[5] Recent work using register data from Denmark by Andersen (2021) documents intergenerational correlations in health indices based on health conditions of around 0.1-0.15. Using U.S. panel data which includes about 8,000 individuals, recent papers by Halliday et al. (2019, 2021)—focusing on intergenerational transmission of self-reported health status or health indices—find somewhat larger correlations of around 0.26.

[6] A number of early papers document the intergenerational transmission of lifespan using relative mobility measures, beginning as early as 1899 with the classic work by Beeton and Pearson (1899, 1901). 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 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. 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 (e.g. data on aristocrats) that are not known to represent any specific population. A notable exception is the recent paper by Kaplanis et al. (2018) which uses a non-representative sample of 130,000 individuals. They find low correlations in lifespan (0.12) and no trend from 1650 to 1850. Other notable exceptions are Ruby et al. (2018) who use data from Ancestry family trees to examine correlations in lifespan between siblings and cousins. They find a large role for assortative mating in the estimation of the heritability of lifespan and find a substantial role for environmental influencing factors. 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 A.1.

Halliday et al. (2021) and Fletcher et al. (2023) find using health measures in relatively small samples: SES measures have higher persistence than measures of health, and correlations across persistence measures of SES and of health are low. But we find that persistence in lifespan is smaller than persistence in health measures estimated in contemporary samples. The contrast in our findings suggests that quality of life, measured by SES and health, is more strongly transmitted across generations than quantity of life.

II. Data and Descriptive Statistics

We exploit rich lifespan data that is available on FamilySearch, a public wiki-style genealogical platform that includes over 13 million registered users and has profiles for over 1.2 billion deceased individuals, making it one of the largest collections of publicly available international lifespan data. 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 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. Most of these profiles are created by people who are doing research on their own ancestors. When individuals are doing research on the same ancestor, they all contribute to the same profile and have the ability to correct any errors that are made by others. The platform also works to enhance and verify records. 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).

Figure A.0 outlines the steps in the data construction and how we arrive at our final estimation sample. We use a base sample of individuals from the full-count US censuses for 1900, 1910, and 1920.[7] We focus 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, some individuals appear in multiple

[7] 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.

census records. We find personal identifiers (PIDs) for 133.3 million person-year observations in our base sample. These person-year observations correspond to 86.6 million unique individuals. Of those 86.6 million unique individuals, 39 million have a death date in their profile.

We also gather the PIDs for 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. Unfortunately, we do not have cause of death information.

a. Data Quality

To examine the accuracy and representativeness of the lifespan data derived from the FamilySearch Family Tree, we use the Social Security Administration (SSA) cohort life tables by year of birth and sex, which are available for the 1900 birth cohort onwards.[8] These national estimates are produced from state reports and try to account for missing populations: not all states are in the vital registration system (e.g. in 1900 only 10 states were in the death registration system, by 1920 there were 36), and there is under-registration in registration states (Hetzel, 1997). Although it is impossible to ultimately determine whether these data are representative of the nation, they constitute the gold standard.

The top panel of Figure A.1 plots the raw lifespan histogram for our sample and for the SSA data, from age zero. We focus on the 1900 birth cohort because the SSA estimates are most accurate for census years, during which populations are observed instead of estimated. Compared to the SSA data, our sample greatly underrepresents individuals who died during childhood. 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 bottom panel), the distribution of lifespan in our data is much closer to the distribution in the SSA data. Conditional on surviving to 25, the bottom panel of Figure A.1 shows that the distribution of the age at death is shifted

[8] Available at https://www.ssa.gov/oact/HistEst/CohLifeTables/2020/CohLifeTables2020.html

right in our sample, so that the mean age at death is slightly higher in our sample, but the shape of the distribution is quite similar.[9]

For the 1900 birth cohort, the expected age at death conditional on survival to age 25 for males in our data is 70, whereas it is 68 in the SSA data. For females it is 76 in our data and 74 in the SSA data. The differences are smaller for the 1910 birth cohorts, for whom the gap is about a year.[10] Part of this difference is likely because our final sample underrepresents African Americans and immigrants, two groups who had higher than average mortality rates (Hacker 2010, Fogel 1986). In addition, the SSA tables were limited to data from states that had Vital Registration Systems in place. They re-weight their sample to make it representative of the US, but the weighting might not be sufficient. Our data, by contrast, includes all states.[11]

b. Our Sample

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, to make the analysis comparable to what is done for economic mobility (which typically restricts attention to age 25 or later when economic outcomes are observed), and to address the issue of missing deaths among children, we restrict our main sample to parents and children who lived to at least 25. Thus, our study investigates adult lifespan, which we define as the age at death among those who survived to age 25.[12]

Our final sample includes 26.1 million child-parent pairs.[13] This sample includes at least 30% of the target population (individuals in the 1900-1920 censuses who were born between

[9] This is also clear from looking at survival rates. Using Kaplan-Meier survival methods, in Figure A.2 we plot the survival functions of the 1900 and 1910 cohorts by gender conditional on survival to age 25 for both the SSA data and our sample data.

[10] Table A.2 shows more details of the differences between our data and the SSA data for the 1900 and 1910 cohorts as well as the gender breakdown of our sample relative to the SSA sample by age for these cohorts.

[11] 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.

[12] Previous work in demography usually restricts attention even further to ages 30, 40 or even 60. As we show later, our conclusions are robust to restricting the sample to older ages. This is most consistent with the prior literature in the study of heritability of lifespan and in studies of the intergenerational correlation of SES measures—almost all previous work conditions on surviving to adulthood. (See Appendix Table A.1.)

[13] We also remove from our sample a handful of individuals who have a lifespan greater than 110 since, for the cohorts we consider, these longer lifespans are likely a result of measurement or reporting error. Because there is

1880 and 1920, survived to age 25 and whose parents also survived to age 25).[14] These data have several advantages over previous datasets constructed to study lifespan transmission. First because we sample our population from the Censuses, we can establish the representativeness of the data – genealogies alone, which are the primary source of demographic work, are typically unrepresentative. Second because we use family trees to identify families, we are not limited to studying families living together at a point in time, as would be the case if we tracked individuals from a given census. Third because the family tree has data from a very large number of sources, our age at death information is more complete, available for all years and not subject to truncation issues, which is an important limitation in other datasets that obtain age at death from national databases such as the Death Mortality Files or the Numident, that only cover deaths occurring within a limited period (eg 1988-2005). Finally, because our samples are very large, we can study heterogeneity.

c. Summary Statistics

Table 1 presents summary statistics for our analysis sample. Our data includes information from the original census record, including race, place of birth, and place of residence. We also have the total number of siblings and the individual's birth order which come from the Family Tree. These summary statistics make clear that our sample, while large, is not representative of all groups in the population: 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 (also see Table A.3).[15] We do, however, have good representation of women, who are typically under-studied in analyses of intergenerational correlations (Hollingsworth 1976). Our data is also skewed towards the Midwest and under-represents the North relative to the full census data.

Although non-representativeness remains an important issue in our study, our sample is substantially larger than samples used in previous research, including the cohort tables from the

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.

[14] There are 89M individuals born from 1880 to 1920 observed in the 1900, 1910 and 1920 census (see figure A.0), when they are ages 0-20. A lower bound of the fraction of the target population we have in our data is 30% (26.1/89). However, many of these individuals in the census will not survive to age 25. For example, according to the SSA cohort tables, 26% of men born in 1900 died prior to age 25. Because there are no cohort tables for the cohorts born 1880-1900 it is not possible to compute the exact fraction of the targeted population we include.

[15] It is well known that genealogies underrepresent Blacks and immigrants, as well as children (Pope 1992).

SSA which include only a subset of states. Whenever possible, we compare our means to SSA data. We also present results separately by race, nativity, gender, cohort, and state of birth--which is available for all native-born individuals. To our knowledge no estimates of lifespan and its transmission exist for many of these subgroup (e.g. Blacks or immigrants).

The average man in the sample (conditional on surviving to age 25) lived to age 70.2 and the average woman lived 6 years longer to age 76.1. This 6-year gender gap is consistent with previous research based on period data (Goldin and Lleras-Muney 2019), and is consistent with the gender gap in the SSA data, which is 6.2, 6.97 and 6.35 for the 1900, 1910 and 1920 birth cohorts surviving to 25.[16] Note that, since we restrict to individuals who lived to age 25, the year of death ranges from 1905 to today, and its 10-90 percentiles are 1946 and 2000, so most individuals in our data (90%) lived beyond the 1918 flu pandemic and WWII and died in the second half of the 20th century (see Figure A.3).

Figure 1 shows the trends in cohort life expectancy at age 25 for cohorts born from 1880 through 1920. For cohorts born before 1890, adult lifespan *fell* for men and was stagnant for women.[17] This 19th century decline for men is consistent with the observed decline in heights among US men for birth cohorts born 1830 to 1890, and with the increases in period mortality observed in the middle of the 19th century. This decline is usually ascribed to the poor sanitary and pollution conditions that accompanied urbanization and industrialization (Pope 1992; also see Costa 2015 for a review).[18]

Starting in 1890 however, there was a steady increase in lifespan for both sexes. Among those surviving to age 25, the age at death rose from 73.5 to 78.1 for women and hovered

[16] https://www.ssa.gov/oact/HistEst/CohLifeTables/2020/CohLifeTables2020.html

[17] 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. However there appears to be consensus based on period data that life expectancy at age 10 (or 20) fell sometime in the 19th century and then started increasing (Costa 2015, Hacker 2010). Although there is no consensus on the exact timing of the reversal most demographers agree--based on period data--that life expectancy rose after 1880. However, cohort data and period data can show divergent trends as noted in the text. The only cohort data we are aware of comes from Pope (1992) who reports declines in adult cohort life expectancy for cohorts born 1760 to 1830 and an increase thereafter, with cohort life expectancy for the 1880-1889 cohorts almost identical to those born in 1760-1769.

[18] The decline in male life expectancy for cohorts born in the late 19th century is also observed in France: cohort life expectancy of French men fell for men born in the late 1870s and until the end of the century, while period life expectancy was constant (see Figure A.4). The extent to which this decline is due to WWI and WWII is unclear but it cannot be uniquely ascribed to wars: if wars were the main determinants, then life expectancy would be falling for cohorts born between 1888 and 1895 and then rising, and then falling again for cohorts born between 1905 and 1923 (based on service rates in the US for these cohorts, see Tan 2020). In a companion paper, we show that the decline in lifespan for men is in part due to their greater survival to age 25 (Beltrán-Sánchez et al. 2024).

between 69.5 to 71.5 for men. There are many possible contributors to these increases. During the second half of the 20th century, when most of the individuals in our cohorts died, the main driver of improvements in life expectancy was technological innovation, particularly in the prevention and treatment of cardiovascular disease (see Beltrán-Sánchez et al. 2024, Cutler et al. 2006).

The second notable trend is that the female advantage grew substantially. This occurred in part because, for men, adult life expectancy was falling in the 1880s and only started rising somewhat in the 1890s. But the gap also grew because, after 1900, the increase in the age at death is larger among women. The rise of the female advantage in lifespan has been documented before using period data which marks 1900 as the year in which this advantage first emerges (Preston and Wang, 2006; Beltrán-Sánchez et al. 2015; Cullen et al. 2016). Our cohort data shows that in fact the female advantage was already large at the turn of the century, at least among adults, suggesting that maternal mortality declines were not the only source of the female advantage, as has been noted before.[19]

Finally, the figure also shows the time series patterns using the SSA data, beginning in 1900, for comparison. As noted above, cohort series for the nation are available from the SSA beginning in 1900. The increases in adult lifespan observed in the SSA data are very similar to those observed in our data, although the increases in adult lifespan are somewhat steeper in the SSA data. Thus, our data shows similar levels and qualitative trends to the SSA data. This suggests that our data does not suffer from large systematic biases and is broadly representative of the US population, despite the under-representation of smaller groups.[20]

III. Absolute mobility across generations

We first investigate the transmission of lifespan from parents to children by plotting the joint distribution of lifespan across generations (Figure 2). The children's adult lifespan is on the y-axis and the parents' adult lifespan is on the x-axis. Lighter colors indicate more density at that

[19] These findings suggest that the female advantage is not uniquely driven by the medical developments in the 1930s and 1940s that lowered maternal mortality (Jayachandran et al. 2010). Instead, declines in fertility, improvements in nutrition (as argued by Fogel 1986, 2004), and beneficial effects of lower exposure to infectious diseases in adulthood likely contributed to these lifespan gains (Goldin and Lleras-Muney 2019). Other factors may have also mattered. There were many substantial social and economic changes in the status of women in the 20th century. See Beltrán-Sánchez et al. (2024) for a longer discussion.

[20] Also notable is that our data does not appear to systematically overstate lifespan, a common observation in demographic studies (Hacker 2010).

location. From these joint distributions, we see a number of important patterns. First, this joint distribution is not normally distributed: for both parents and adults there is a long tail of individuals who die relatively young and the distribution is compressed at the top. Second, while there is a positive correlation between parents' and children's adult lifespan, there is substantial variance.[21]

These figures already suggest that parental adult lifespan is only weakly predictive of child adult lifespan. Indeed, some simple computations confirm what we see in the heat map. For example, the chances that a child will live to be the same age minus/plus 1/2/5 year(s) as their same sex parent are only 6%, 11% and 24%.[22] These shares do not vary across cohorts. So the majority of people live substantially longer or shorter lives than their parents.

Next, we investigate absolute mobility by assessing whether children live longer than their parents. One measure of absolute mobility in lifespan is simply the fraction of children who live longer than their parents, which corresponds to the area of the distribution above the 45-degree line in Figure 2. Given the rise in adult lifespan across cohorts that we observe during this period, we might expect that absolute mobility will increase as well. However, this need not be the case. If all increases in lifespan accrue to children who would already be living longer than their parents, then this measure of mobility will remain constant. There is in fact no mechanical relationship between increases in life expectancy and our main measure of absolute mobility.[23]

Figure 2 documents that absolute mobility in lifespan is substantially higher for women than men: most of the joint female distribution is above the 45-degree line, but most of the joint male distribution lies below the 45-degree line. Indeed roughly 57% of women live longer than their mothers, but only 47% of men live longer than their fathers, a substantial 10 percentage point gap. While we are unaware of similar estimates in the literature, we do know that these

[21] For example, consider men whose father died at age 77 (the modal age at death among fathers in the sample). Although the median age at death among the sons of these fathers is 76, the IQR is 64 to 85. Thus, 50 percent of men will die at substantially younger or older ages than their fathers. Although the variance is lower for women, the IQR range is still large.

[22] If a father died at age 80, we code the son as dying within one year of the father if he died at 79, 80 or 81; within 2 years if he lived to 78-82 and 5 years if he lived to 75-85.

[23] Our aggregate absolute mobility measure could fall even when lifespan is increasing if the gains in lifespan are unequally distributed. For example, suppose one group of children go to war and die young while another group benefits from medical advances. Lifespan in the first group will fall, and they will live shorter lives than their parents. Lifespan in the second group will rise, and they will live longer than their parents. If the first group is larger, overall absolute mobility will fall. But if the gains in years lived in the second group is large, then average age at death might in fact increase.

measures for men are similar to estimates of absolute mobility in income for the 1880-1920 birth cohorts in the 40-50% range, though these estimates are likely subject to substantial error since they are mostly based on occupation rather than income (Song et al. 2020; measures for women are not available).[24]

Our absolute mobility measure does not quantify *how much* better off children are than their parents among those living longer lives, or *how much* worse off are those who died younger than their parents. Figure A.5 shows the distribution of the difference in the ages at death of children and their parent (child-parent). The mean gap is 3.8 for women and -1.4 for men. Thus, the increases in lifespan for women relative to their mothers are quite substantial, and the declines for men relative to their fathers are less so. However, these distributions are symmetric and have large standard deviations, of around 20 years of life for both sexes, suggesting that it is hard to predict the age of death of the child based on the age at death of the parent.

One might expect absolute mobility to differ depending on who benefits most from advances in public health and medical technologies or other policies like health insurance. We investigate heterogeneity in absolute mobility in panel A of Figure 3. We see that, among more highly educated families (as proxied by whether the child has 9 years of schooling or more), the probability that the child outlives the parent is approximately .05 *lower* than the same probability among less highly educated families. Thus, low SES individuals gained more mobility as a result of lifespan increases, particularly among women. These results are consistent with findings that technologies such as sulfa drugs and antibiotics lowered socio-economic disparities in health by improving the health of low SES individuals the most (Jayachandran et al. 2010; Alsan et al. 2021). These developments were particularly important in reducing maternal mortality, an important cause of death among young women before 1936.

We also find significant racial differences, with white absolute mobility higher than that of non-whites, particularly for women. Outside of sex, SES, and race, there are no other meaningful (and consistent) differences across groups, defined by their immigrant status, the immigrant status of the parents, family size, birth order or state of birth, although there is a bit more variation among men than women for all subgroups. It is notable that although there are very large differences in adult lifespan by state of birth (the gap in average lifespan in 1880

[24] Absolute income mobility was substantially larger for cohorts born in 1940 and later, ranging from 90% for children born in 1940 to 50% for children born in the 1980s (Chetty et al. 2017).

between the longest-lived and the shorted lived state was roughly 6 years for women and 8 years for men), the absolute mobility measures are quite similar across geographies.

Figure 4 shows trends in absolute mobility. For the 1880-1900 birth cohorts, absolute mobility was increasing for women but decreasing for men. Around the 1900 birth cohort absolute mobility started to fall for women and started to increase for men.

The story for men is straightforward: the trend in absolute mobility follows the trend in average lifespan. The decline in absolute mobility is consistent with the decline in average lifespan for the cohorts born 1880 to 1900. Then absolute mobility rises as the average lifespan rises (See Figures A6). The story for women is more complicated. Absolute mobility first rises for women as lifespan is rising, but then beginning around 1900, absolute mobility falls even though the average female lifespan is rising for the 1900-1920 cohorts. Why? A detailed examination of the heat maps for the 1900 and 1920 birth cohorts (Figures A7) provides an explanation. There is a small group of daughters born in 1920 living substantially longer lives: they are no longer dying in their 20s and 30s. This might be the result of declining maternal mortality for these later cohorts: maternal mortality started falling significantly only after 1935 and would thus only affect cohorts born after around 1900. These increases explain why the average lifespan is rising. But there is also a relatively large group in 1920 dying in their 70s, living somewhat shorter lives than those born in 1900 (and thus living shorter lives than their mothers). These trends highlight important sex differences in both lifespan and its transmission.[25] They also highlight the importance of observing these joint distributions—and not just average lifespan—to predict trends in absolute mobility.

In sum, measures of absolute lifespan mobility show that mobility was substantially larger for women than men and for individuals who obtained low levels of education. It also varied significantly from 1880 to 1920, rising and then falling for women, and falling and then rising somewhat for men. One might have hypothesized that the introduction of social insurance

[25] Although male absolute mobility is a bit higher for the 1920 cohort than for the 1880 cohort, this gain obscures important differences in absolute mobility by parental age. Figure A.8 shows the probability that the son lives longer than his father for the early (1880) and late (1920) cohorts. Men born in 1880 were *more* likely to outlive fathers who lived short lives (to age 60 or less) than in the 1920 cohort. This is likely due to the 1920s cohort exposure to WWII, which killed a large number of men born around 1915 and 1925. However, this is the opposite of what we see among longer-lived fathers; for fathers who lived longer than 60 years, sons in the later cohorts were more likely to outlive their fathers than sons in the earlier cohort. Because this second effects dominates, on net the mobility is larger for the 1920 cohort. Among women who did not die in large numbers in WWII, we see a similar pattern, although for mothers who lived longer than 40 years, daughters in the later cohorts were more likely to outlive their mothers than daughters in the earlier cohort.

in the second part of the 20th century (including pensions and public health insurance), would have increased health mobility for more recent cohorts. This is somewhat true for men but not for women.

IV. Relative mobility

a. Relative Mobility Measures: Levels and Trends

Most research focuses on relative (rather than absolute) mobility, namely how individuals fare relative to others in their cohort. One way of summarizing the data to study mobility is to compute transition matrices, which show the probability that a child born to a parent in a given quintile (of the parental distribution of lifespan) ends up in a given quintile (of the child distribution of lifespan).

Table 2 shows that the transition matrices for both women and men are close to what we would expect if there were perfect mobility: the diagonal elements are somewhat larger than 20% but not much, and the off-diagonal elements are not substantially lower than 20%. Thus, there is substantial upwards and lower mobility. For example, 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 20% of parents ends up in the top 20% of the income distribution is only 10% (Chetty et al. 2014). The probability of living a long life (top quintile) among those whose parents lived a long life (also in the top quintile) is 25% for both fathers/sons and mothers/daughters, suggesting that long-lived parents have children who are likely to also be long-lived consistent with van den Berg et al. (2019).[26] This is the largest deviation from the expected 20%, but even this deviation—the largest in our sample—still suggests relatively low persistence, at least compared to what has been documented for income.[27] Overall relative mobility measures show high upward mobility, particularly for those born in to short lived parents.

When we look across time, we see that relative mobility appears to be very stable across cohorts (Figure 5). Upward mobility (the fraction of kids born to parents in the bottom 20% of

[26] This is consistent with findings in the literature using siblings. For example, Perls et al. (2002) and Schoenmaker et al. (2006) show that the survival rates of siblings of long-lived individuals are higher than that of the population.
[27] For example, Chetty et al. (2014) find that the share of children born to parents in the top quintile of income that also end up in the top quintile of income is 36.5%.

adult lifespan who make it to the top 20th percentile of the distribution of adult lifespan in their cohort) did not change much for women, remaining at roughly 18% throughout the period. Downward mobility (the share from the top 20% that ends up in the bottom 20%) also stayed constant at around 17% for both men and women. Persistence at the top (the share that stays in the top quintile) was around 25-27% for both sexes. Persistence at the bottom (the share that stayed in the bottom quintile) remained around 22% for both sexes.

Altogether, these results show that absolute and relative measures of mobility give very different pictures of mobility. This is also true when looking at income; for example, Chetty et al. (2017) show that absolute mobility fell for cohorts born 1940 to 1980 while relative measures remained very stable. This highlights the importance of looking at both.

b. Relative Persistence

We next follow the economics and biology literatures and compute measures of persistence, rather than measures of mobility. A commonly used measure of persistence comes from regressing sons' outcomes on fathers' outcomes, which estimates the slope of the joint distribution of lifespan in levels, as shown in Figure 2. As it is obvious from the heat maps, the slopes for both of these relationships appear small.

Before proceeding note that lifespan increases across cohorts have no immediate or mechanical implications for intergenerational correlations or other measures of relative persistence such as regression coefficients. If all children live 20 years longer than their parents (shifting the distribution up), then the correlation in lifespan across generations will remain stable.[28]

Our main specification relates the lifespan of the child (L_i^C) to the lifespan of the parent (L_i^P) by estimating the following equation:

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

where L_i, our main variable of interest, is lifespan in years for individual i and L_i^P is the lifespan of one of the parents, conditional on living to age 25 for both parent and child.[29] We also present results in logs and percentiles. In some specifications, we also allow L_i^P to refer to the average of

[28] Conversely, life expectancy could be stable, but the intergenerational persistence in lifespan (IGPL) could change if the joint distribution changes shape.
[29] We later test the sensitivity of our conclusions to this restriction.

the parents' lifespan. The standard errors are clustered at the family level, since our sample can include multiple children from the same parents.

Our preferred estimates include a parsimonious set of controls, *X*, including indicators for the cohort of the child and the cohort of the parents to account for secular trends in adult lifespan. In some specifications, we also include state of birth fixed effects to proxy for environmental factors, as well as controls for race and immigrant status. We find the relationship remarkably robust to the inclusion of additional controls.

The coefficient of interest is β_1, which we refer to as the intergenerational persistence in longevity (IGPL). In levels, β_1 represents the average increase in lifespan associated with a one-year increase in the lifespan of the parent. In logs, β_1 represents the intergenerational elasticity of lifespan. In percentiles, β_1 represents the Spearman correlation in lifespan. In these specifications, regardless of functional form, the coefficient on lifespan, β_1, will incorporate the influence of parental genetics, assortative mating between the parents as well as socio-economic influences and common environmental factors that affect both parents and children's adult lifespan (see Appendix 1 for a simplified model of the intergenerational correlation in lifespan).[30] As is typical in this literature, we are estimating associations and not the causal effect of exogenously changing parental lifespan.

The raw coefficient from a regression of the lifespan of a son on the lifespan of his father (without other controls) is 0.089 (Table 3, Column 1), while that for a son and his mother is 0.062, consistent with the predictions of a simple model where there is a sex-specific component to lifespan (Appendix 1). 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. 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 lifespan. Similarly, when we look at daughters, we see that the daughter/father coefficient is 0.075 and the daughter/mother one is 0.081. When we relate the lifespan of a daughter to that of the average of her parents' lifespan, the coefficient is 0.15.[31] We obtain nearly identical coefficients when we include parent and child birth-year fixed effects (Column 2), state-of birth fixed effects for both parents and children (Column 3), or controls for race and birth order

[30] It can also incorporate the causal impact of parental lifespan on child lifespan if there is one.
[31] The correlation between father and mother adult lifespan is quite low, at only .04.

(Column 4). While these characteristics have significant effects on child's lifespan, the coefficients on parental lifespan are unchanged.

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 presented earlier (Column 1), a rank-rank specification, where percentiles are calculated within birth cohort (Column 2), and a log-log specification (Column 3). Although the coefficient estimates are a bit lower in the log-log specification, they are very similar.[32] Thus regardless of which specification we use, we find that the relative persistence is always low.

To check that our linear specification is appropriate, we visually plot the relationship between the lifespans of parents and children (Figure A.9). 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 very top.

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 Table A.1). Studies focused on the role of genes in the intergenerational persistence of lifespan 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. Figure A.10 shows these correlations are remarkably robust to the choice of cutoff ages (which we allow to vary from 0 to 65), as they remain low throughout. These results suggest that the genetic contributions to lifespan are small, an issue we return to later.

While the overall correlation is low, it may be the case that different groups experience different levels of persistence. For example, if some groups benefit more from new medical technology or investments in public health, we may see differences in the IGPL. However, when we examine heterogeneity by parent's characteristics, child characteristics, and place of birth

[32] Table A.4 shows raw correlations instead of regression coefficients, since these are often reported in other studies. Raw correlations in lifespan 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.

(Figure 3b), we see surprisingly little heterogeneity. Even when we do observe differences, the range of the IGPL is always between 0.06 and 0.10, suggesting low persistence for all groups. This is true whether we look at levels or rank correlations. The most interesting finding here is again by education: more highly educated families (as proxied by children with 9 years of schooling or more) experience more persistence in lifespan—both fathers/sons and mothers/daughters—than children in less educated families, although, again, the point estimates are all relatively small. Similarly, immigrants seem to have slightly less persistence than natives, although this difference seems to disappear by the second generation. There is also some variation across states, although again persistence is low throughout.

c. Time Trends in Relative Persistence

Figure 6 shows the IGPL in levels 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 these increases are large in a relative sense, the magnitudes remain low and lower than estimates for education or income.

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 from around 0.07 to about 0.1, a smaller increase than when we compute this trend in levels. Moreover, rank-rank specifications tend to show very little difference in the evolution of the IGPL across sexes, whereas we see different patterns for men and women when we examine relative persistence in levels as well as when we look at measures of absolute mobility, highlighting again that different measures of mobility can tell a very different story, as noted by Deutscher and Mazumder (2023). Interestingly, the rank-rank correlation in occupational status also remained relatively constant for men born 1880-1920 (Song et al. 2020, Ward 2023, Buckles et al. 2023) and women born in this same period (Buckles et al. 2023). Finally, there is more persistence over time for both sexes in all our relative measures, so the expansions of social insurance and health insurance did not increase relative mobility for more recent cohorts—if anything they lowered it.

Overall, we find little relative persistence in lifespan across generations: although longer-lived parents have children who are themselves longer-lived on average; the effect is small. This conclusion does not depend on how we estimate persistence and is not limited to any particular time or group in the population. This stands in contrast with results for occupational status, which show much higher persistence for these cohorts: the rank-rank correlation is estimated to be somewhere between 0.3 (Song et al. 2020) and 0.7 (Ward 2023, Buckles et al. 2023). Our estimates of persistence are also lower than contemporary estimates of persistence in health, which are around 0.3 (e.g. Halliday et al 2021.)

V. Comparing Mobility in Lifespan and in Socio-Economic Status

Our results so far suggest there is significantly more mobility in lifespan than in other economic outcomes, such as education and income. However, this conclusion is based on a comparison to the literature, which has used a variety of different time periods and data selection criteria. As a result, it is unclear whether these differences are due to true differences in persistence or differences in samples. While we would like to compare intergenerational persistence in lifespan to that of education and income, we do not have sufficient data on the education and income of the parents in our sample; instead, we look at sibling correlations, which provide an upper bound on intergenerational persistence.

To compare estimates of the intergenerational persistence of lifespan with persistence in economic measures for the same population, we construct a subsample of sibling pairs for whom we observe lifespan and whose education, occupation, and income were observed in the 1940 census.[33] Our sibling sample is observably very similar to the full analysis sample (Table A.5). We divide our analysis here to compare brothers with brothers, sisters with sisters, and sisters with brothers (Table 5).

We regress each individual's outcome on the outcome of their same-sex sibling or opposite-sex sibling, averaging across siblings if they have more than one. We estimate sibling coefficients for adult lifespan, education, and earnings (both individual and household) using a constant sample (Columns 1 – 4) and include birth cohort fixed effects for each sibling. Standard errors are clustered by family. Because many women do not have positive earnings in the 1940

[33] 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.

census, we also consider a less restricted sample, where we include all individuals with non-zero measures of education and lifespan (but no restriction on earnings) as a robustness check (Columns 5 and 6).

Several conclusions emerge from these sibling correlations. First, the coefficients for lifespan among siblings are larger than our parent-child intergenerational correlations: for example, the brother-brother coefficient is 0.13 and the father-son coefficient is 0.08 for this same sample (shown in the last column). This is unsurprising. If the parents and children grew up in the same environment, in a simple model of longevity one would expect the sibling correlation and parent/child correlations to be the same (see Appendix). However, to the extent that siblings grow up in a more similar environment than parents and children—as siblings share the same parents, home, and neighborhood—sibling correlations will exceed the measure of intergenerational persistence. [34] This is particularly likely to be true during the time period under study, given the large changes in economic and disease conditions occurring throughout the lifetime of these cohorts.

Second, sex plays a similar role among siblings as with parents and children, as would be expected if genetic, social, and economic factors differ substantially by sex. Similar to the child-parent coefficients, we see a larger coefficient for male siblings' adult lifespan (0.13) than for female siblings' (0.11). We also see a rather weak relationship between brothers' and sisters' lifespan (0.04).

Third, the correlations in SES we document are consistent with those in the existing literature, albeit a bit larger, 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 instead of regression coefficients, we get correlations of 0.55 for brothers, 0.59 for sisters, and 0.53 for mixed-sex siblings (Table A.6). This is consistent with the literature, which calculated estimates for more recent cohorts. For example, for men, Solon et al. (1991) estimate that the sibling correlation in education is 0.45 for cohorts born in 1944-1958, and Björklund and Jäntti (2020) estimate 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

[34] See Björklund and Jäntti (2012) for only one example of this. 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.

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 and 0.17 for females, comparable to those estimated in the literature, which generally range from 0.1 to 0.4 using single-year earnings.[35] We can also compare our sibling estimates of education with what we would obtain in our sample if we looked at the highly selected subset of children whose parents are also observed in the 1940 census and for whom we also have education. The intergenerational persistence in education ranges from 0.4 to 0.5 (Table A.7) in line with modern estimates for the US (Black and Devereux 2011).

One possibly puzzling finding is the negative sister/brother coefficient on income; however, this is likely driven by the fact that women from wealthier families worked less than women from poorer families, which could lead to a negative relationship between brother's income and sister's income. To address this, Column 4 shows coefficients for the relationship in household (rather than individual) income. As expected, 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 this fixed sample of individuals, we see much larger coefficients (and correlations) for education or income than for lifespan. The coefficients for household income are nearly three times as large as those for lifespan: 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 lifespan. Table A.7 shows that, in the highly selected sample of parents and children with both education and lifespan, the conclusions are similar, with persistence in education of around 0.45 and persistence in lifespan of about 0.15. These findings confirm our earlier observations that persistence in SES is substantially larger than persistence in lifespan.

a. The Relationship Between Persistence in SES and Persistence in Lifespan

Finally, we examine whether persistence in SES and persistence in adult lifespan are related across states and over time. Although lower in levels, one might expect that

[35] See Solon (1999). Correlations are in fact higher when more permanent measures of income are computed (Eshaghnia et al. 2023).

intergenerational correlations in adult lifespan are driven by the same factors that make SES persistent. In areas or periods with more equality of resources—which may lower intergenerational persistence—society likely spent 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 more persistent from one generation to the next, health would also be more persistent as well. Intergenerational correlations in education, income and health might also be high in the same locations or periods because education, income, and adult health are highly correlated.

We use our sibling sample to estimate sibling correlations in education and sibling correlations in lifespan by birth state or by cohort to investigate this. Because our measures of education are taken from the 1940 census, cohorts born in 1880 will need to live to age 60 to be observed in the 1940 census. As a result, we restrict our sample to siblings who survive until at least 60 years old for this analysis so that all cohorts are conditioned to live until the same age and can be compared.[36] In Figure 7, we plot the coefficient on lifespan (y-axis) across states (or across cohorts) against the coefficient on education (x-axis).

Surprisingly, when we look across states (on the right-hand side of the figure), we see that correlations in SES and correlations in adult lifespan are not systematically related; places where the correlation in education is large are *not* the same places where the correlation in lifespan is high. A simple regression finds that the slopes of all of these regression coefficients are very small and, in most cases, statistically insignificant. Results using our parent child sample to estimate persistence across generations in education and lifespan are very similar. [37]

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 are substantially smaller than correlations in all other SES measures. Moreover, this persistence of adult lifespan and of socioeconomic status are unrelated. This highlights that families and communities that succeed

[36] Given that we observed similar measures of persistence when we restricted to those living until age 60 as we did when we restricted to those living until age 25, this is unlikely to affect our conclusions. However, we have verified that our across-state analysis is similar when we examine the sample of those living until age 25 as well.
[37] Since sibling correlations are not identical to intergenerational correlations, we repeat this exercise using the highly selected subsample of parents and children with education and lifespan. To make coefficients comparable across cohorts we restrict attention further only to individuals that survive to age 60. Then we plot these estimates to see if they are correlated. In Figure A.11, we plot the coefficient in lifespan (y-axis) across states (or across cohorts) against the coefficient in education (x-axis). Despite the fact this sample is highly selected, the results are very similar, confirming these association are low.

in increasing mobility in one domain do not necessarily do so in others. This is consistent with recent research using contemporary survey data that documents that places that have higher mobility in SES are not necessarily those with higher mobility in health.[38]

The fact that the persistence of SES and the persistence of adult lifespan are independent of each other implies that mobility in well-being may be larger than economic measures alone would suggest. Thus, measures of persistence in SES paint an incomplete picture of persistence in well-being.

VI-Why is Mobility So High?

As noted in the introduction, our finding that persistence in adult lifespan is low is not new. Table A.1 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 the present day. Figure 8 plots the estimates from a few of these studies and compares them directly to ours (we estimate the identical specification with our data whenever possible). 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. The limited variation across time, space, and other characteristics that we documented in the previous sections provides an explanation for why the results across the literature are so consistent: the persistence in adult lifespan is low and there is limited heterogeneity (in an absolute sense) in its magnitude.

These results might seem unintuitive. There are at least three reasons why, based on previous work, one might expect lifespan to be highly persistent across generations. First, it is well known that there are important genetic determinants of health, and genes are transmitted from parents to children. Second, SES is an important determinant of lifespan and SES is highly persistent across generations.[39] Third, behaviors that determine health and SES are also passed

[38] Fletcher and Jajtner (2021) investigate the same question using a contemporary sample of about 16,000 children from the AddHealth survey and mobility measures for many outcomes, including measures of health (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 a sample of about 8,000 individuals in the PSID and compare mobility in self-reported health and in incomes. They find that the two measures are only weakly related. Thus, mobility appears to be an outcome specific process, that depends on inputs that are outcome specific. Moreover, this conclusion appears to hold regardless of the health measure one uses, and the period examined.

[39] In Table A.8, we show that, in our data, education is a significant predictor of lifespan.

on from parents to children. By passing on genetics, behaviors and SES, the family would, ex ante, appear to be an important determinant of health. On the other hand, measurement error in adult lifespan could lower our estimates of persistence and as a result show a lot of upwards and downward mobility.

a. Bounding the Role of Measurement Error

The first issue we consider is the possibility that our low measures of intergenerational persistence in lifespan are an artifact of lifespan being poorly measured. While this is unlikely, given the consistently low estimates obtained in the existing literature using a variety of samples, we rule out this possibility by re-estimating our intergenerational correlations and sibling correlations on a portion of our sample that includes only pairs in which both individuals (parent/child) have a death record attached to their family tree (Appendix Table A.9). Despite the smaller sample size, the IGPL is basically unchanged. To rule out that this is the result of low quality in the father's death certificate information (which could be low given that in the US vital records were not well kept until the 1930s), we compare sibling correlations instead. We 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 sufficiently large to produce correlations in the range of the SES correlations we observe.

b. Bounding the Contribution of Genes

To better understand the role of genes in the intergenerational correlation in lifespan, we take advantage of the fact that we have data on both siblings and twins, who have varying degrees of genetic connections but who grow up in the same family (and hence the same environment). If genes are an important factor, we would expect the correlation in lifespan to be much higher among twins than among siblings. In contrast, if genes are less important (or unimportant), we would expect the correlation in lifespan to be similar among siblings and twins, as both have similar family environment but varying degrees of genetic linkages.

We have a large sample of twins – more than 100,000 pairs of twins with information on adult lifespan.[40] Unfortunately, we cannot differentiate between identical and fraternal twins, which means our estimates will understate the role of genetics; however, since we know the

[40] We identify twins as siblings who are born to the same parents in the same year and month.

share of same-sex twins who are identical in the population, we can adjust our estimates accordingly.

Table 6 shows the results. Brother-brother lifespan coefficients are larger among twins (0.18) than among siblings (0.13), as expected. Similarly, sister-sister lifespan coefficients among twins (0.16) are larger than female sibling correlations (0.10). This is true even though the intergenerational persistence in lifespan with the father is almost the same among siblings and twins.

However, as noted above, we are understating the role of genetics in these estimates, as some of the children classified as twins are actually fraternal—and not identical—twins. Since approximately 50 percent of same-sex twins were identical during this period (Jeanneret and MacMahon 1962), we can use this fraction, along with the same-sex twin and sibling correlations, to determine the coefficients for identical twins. We calculate that, for identical male twins, the coefficient of the persistence in lifespan is 0.23 and for identical female twins the same coefficient is 0.22, approximately 50 to 70 percent larger than the same-sex sibling coefficients in lifespan. These identical twin correlations can likely be viewed as an upper bound on the heritability of lifespan, since identical twins share the same genetics, while parents and children share only half. In addition, twins also have more similar home environments relative to non-twin siblings. The identical twin correlations in lifespan are still below 0.25, suggesting that the genetic heritability of lifespan is relatively low.

Our conclusion regarding the moderate role for genes in explaining lifespan is consistent with the findings in other studies using twin designs, which find that genes explain about 25% of the variation in lifespan (see review by Dato et al. 2017).[41] However, recent research looking to identify specific genes that affect lifespan has produced estimates that are significantly lower, consistent with the idea that even twin studies overstate the contribution of genes to outcomes.[42]

[41]Closely related is work by Hjelmborg et al. (2006) that uses Danish, Finnish, and Swedish twins born between 1870 and 1910 and finds that genetic influences on lifespan are minimal prior to age 60 but increase thereafter. There is also important new work trying to disentangle the role of nature versus nurture in the intergenerational transmission of lifespan using data on adoptees. Björkegren 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.

[42] The most recent genome-wide association study identified only 12 single nucleotide polymorphisms or SNPs (out of millions of possible SNPs candidates) that affect lifespan (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

Altogether our results confirm genetic influences in determining adult lifespan are modest, at least relative to the influence of genes in the determination of other outcomes.[43] Our upper bound for the contribution of genetics in the determination of lifespan is about 25%, which is much lower than estimates for the genetic contribution to height (68-80%) and in the same range as the contribution of genes to body mass (30%) (Wainschtein et al. 2022) or education (14-18%) (Okbay et al. 2022).

c. Bounding the Contribution of Overall Family Factors and Assessing the Role of Luck

An explanation for the low level of persistence in lifespan is that the stochastic component of lifespan is larger than it is for other outcomes. Indeed, in a theoretical paper on this issue, Vaupel (1988) demonstrates this using a simulation of his frailty model, where frailty is a measure of an individual's health or disease susceptibility which determines their probability of dying. He shows that even if the correlation in (unobserved) frailty between parents and children is exactly equal to 1, the correlation in lifespan can be close to zero if the stochastic component of lifespan is large. This would explain why the persistence in health (and its SES determinants) is higher than the persistence in lifespan we estimate.

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

Table 7 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 an individual's lifespan on all observables, not including parental or siblings' lifespan, then the R-squared of the regression is low, 0.04; when we estimate the same specification for education,

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).

[43]A possible reason why genetic influences are small is that there is little evolutionary pressure to transmit genes that increase longevity beyond reproductive years.

the R-squared of the regression is 0.13 (Panel B).[44] When we include education, income in 1940, and occupation, the individual coefficients on these variables are statistically significant but the R-squared of the regression remains low, in the range of .04. When we include family fixed effects, the R-squared increases substantially for education (to 0.73) but much less so for lifespan (0.38), demonstrating that families have a more limited influence on lifespan than they have on education. Altogether these findings suggest that Vaupel's (1988) hypothesis is consistent with the data: the IGPL is low because the stochastic component of lifespan is large, and larger than for health or SES outcomes.[45]

VII. Conclusion

While there is a robust literature in economics examining intergenerational correlations 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 persistence in lifespan.

We find that absolute mobility in lifespan—the probability that a child lives longer than their same-sex parent—was larger among women, and it moved in opposite directions for men and women for the cohorts we study. When we examine relative mobility, we find that the intergenerational persistence in lifespan (a measure of persistence rather than mobility) is strikingly low in a sample that includes more than 30% of the adult population in the US at the time. Consistent with research in other fields, father-son correlations are in the range of 0.09, and correlations with both parents on the order of 0.14. We also find that persistence is low across subgroups and over time, which provides insight into why earlier, very different studies, often with relatively small convenience samples, come to similar conclusions. The consistency of

[44] These observables include the birth cohort of mother fixed effects, birth cohort of father fixed effects, child cohort fixed effects, place of birth fixed effects, indicators for race, gender, number of siblings, birth order, mother and father immigrant status.

[45] These findings are also consistent with the model of cohort health and mortality recently proposed by Lleras-Muney and Moreau (2022) in which the age at death is stochastic and determined partly by investments, and partly by the accumulation of shocks over the lifetime. These findings suggest that an interesting avenue for future research is to estimate the correlation in the underlying risk (or frailty) that is transmitted across generations; lifespan is clearly a poor proxy.

estimates across studies, samples, and times implies that the results we obtain for the 1880-1920 cohorts may be applicable to cohorts born in more recent times.

However, while persistence is low for all, there is substantial variation in these estimates across subgroups and over time in *relative* terms. While these results suggest there may be notable heterogeneity, future work will need to examine more carefully the extent to which the subgroup samples are more broadly representative. Nevertheless, since our measures of persistence are low for all groups, the results suggest that information about parental ages at death is not particularly informative about the age at death of an individual. Therefore, households and governments cannot rely on this information to make accurate predictions about individual lifespan.

Our data suggest that longevity correlations are low (and much 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. However, the intergenerational transmission of risk—of which age at death is only one realization—might be much higher. In fact, the intergenerational transmission of health, which might be a closer proxy for underlying risk, has been estimated to be substantially higher. Future research could attempt to estimate the intergenerational transmission of risk across individuals directly, a far more challenging enterprise.

Previous studies have focused on relative mobility. However absolute and relative mobility measures give a very different picture of intergenerational mobility. While absolute measures show very different levels and trends by sex and race, the results across groups are strikingly similar when we consider relative mobility measures instead. We conclude, like others have when looking at income, that in the context of health, absolute and relative measures paint very different pictures.

Taking advantage of data on the family structure, we then 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. Importantly, we find this correlation remains low over time, suggesting limited evidence of aggregate effects of health policy or technological and medical advances that increased lifespan on mobility over this time. We also document that sibling correlations in lifespan across states or cohorts are surprisingly unrelated to sibling correlations in education,

suggesting that the factors that determine the IGC in lifespan will be different from those that determine the IGC in education. An important direction for future research will be to understand how different intergenerational measures are related, and further to combine absolute and relative measures of lifespan, health, and SES into a single metric of well-being to better understand its intergenerational transmission.

In the process of conducting this analysis, we also document new patterns in longevity, as our data begins with the cohort born in 1880, 20 years prior to when the SSA data begins. We show that the cohorts we study experienced large changes in lifespan. Most notably, we document that for men, the end of the 19th century was a period associated with declines in lifespan. On the other hand, lifespan rose very substantially for women during this period. Why men's lifespan fell while women's did not is the topic of other work, but these important sex differences in lifespan, combined with the striking differences in the patterns of absolute mobility by sex, highlight the importance of careful consideration of men and women separately.

IIX. References

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez (2021). Intergenerational Mobility of Immigrants in the United States over Two Centuries. *American Economic Review*, 111 (2): 580-608.

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.

Alsan, M., Atella, V., Bhattacharya, J., Conti, V., Mejía-Guevara, I., & Miller, G. (2021). Technological progress and health convergence: the case of penicillin in postwar Italy. *Demography*, *58*(4), 1473-1498.

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.

Arias, E., Xu, J., Tejada-Vera, B., Murphy, S. L., & Bastian, B. (2022). US State life tables, 2020.

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.

Bakhtiari, E. (2022). The missing mortality advantage for European immigrants to the United States in the early twentieth century. *Demography*, *59*(4), 1517-1539.

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

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.,

Beltrán-Sánchez, Hiram, Sandra E. Black, Adriana Lleras-Muney and Joseph Price (2024). The Evolution of Lifespan Among Cohorts Born at the Turn of the 20th Century US. Mimeo.

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 Lifespan 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

Buckles, K., Price, J., Ward, Z., & Wilbert, H. (2023). Family trees and falling apples: historical intergenerational mobility estimates for women and men. National Bureau of Economic Research Working Paper Number 31918.

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., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review*, *104*(5), 141-147.

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.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., & Narang, J. (2017). The fading American dream: Trends in absolute income mobility since 1940. *Science*, *356*(6336), 398-406.

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.

Couillard, Benjamin K., Christopher L. Foote, Kavish Gandhi, Ellen Meara, and Jonathan Skinner. 2021. "Rising Geographic Disparities in US Mortality." *Journal of Economic Perspectives*, 35 (4): 123-46.

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 lifespan: 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 lifespan genes. *Nature Communications*, *10*(1), 1-14.

Deutscher, N., & Mazumder, B. (2023). Measuring intergenerational income mobility: a synthesis of approaches. *Journal of Economic Literature*, *61*(3), 988-1036.

Deryugina, T., & Molitor, D. (2021). The causal effects of place on health and lifespan. *Journal of Economic Perspectives*, *35*(4), 147-170.

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.

Eshaghnia, S., Heckman, J. J., & Landersø, R. (2023). *Maximum Impact Intergenerational Associations* (No. w31093). National Bureau of Economic Research.

Feigenbaum, J., Hoehn-Velasco, L., & Wrigley-Field, E. (2020). Did the Urban Mortality Penalty Disappear? Revisiting the Early Twentieth Century's Urban-Rural Mortality Convergence. Minnesota Population Center Working Paper Series.

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 JM, Engelman M, Johnson NJ, Hakes J, Palloni A. Understanding Geographic Disparities in Mortality. *Demography*. 2023;60(2):351-377. doi:10.1215/00703370-10609710.

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 Lifespan. 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. (2023). Intergenerational health mobility. In *Handbook of labor, human resources and population economics* (pp. 1-22). Cham: Springer International Publishing.

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.

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

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.

Hetzel, A. M. (1997). *US vital statistics system: major activities and developments, 1950-95*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Hjelmborg, J. V., Iachine, I., Skytthe, A., Vaupel, J. W., McGue, M., Koskenvuo, M., ... & Christensen, K. (2006). Genetic influence on human lifespan and lifespan. *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.

Kumar, S., & Nahlen, B. (2023). Intergenerational persistence of health: Evidence from India. *Economics Letters*, *224*, 111023.

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

Jayachandran, S., Lleras-Muney, A., & Smith, K. V. (2010). Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs. American Economic Journal: Applied Economics, 2(2), 118-146.

Jeanneret, O., & MacMahon, B. (1962). Secular changes in rates of multiple births in the United States. *American Journal of Human Genetics*, 14(4), 410-25.

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.

Kasakoff , A. B. and Adams , J. W. 1995 . The effect of migration on ages at vital events: A critique of family reconstitution in historical demography. European Journal of Population , 11 : 119 – 242.

Kemkes-Grottenthaler, A. (2004). Parental effects on offspring lifespan—evidence from 17th to 19th century reproductive histories. *Annals of Human Biology, 31*(2), 139-158.

Kerber, R. A., Obrien, E., Smith, K. R., & Cawthon, R. M. (2001). Familial Excess Lifespan 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 lifespan: 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.

Ljungquist, B., Berg, S., Lanke, J., Mcclearn, G., Pedersen, N. (1998). The effect of genetic factors for lifespan: a comparison of identical and fraternal twins in the Swedish Twin Registry. *The journals of gerontology*, *53*(6), 441-446.

Lleras-Muney, A. (2022). Education and income gradients in lifespan: The role of policy. *Canadian Journal of Economics/Revue canadienne d'économique*, *55*(1), 5-37.

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

Lu, F., & Vogl, T. (2023). Intergenerational persistence in child mortality. *American Economic Review: Insights*, 5(1), 93-109.

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.

Manski, C. F. (2011). Genes, eyeglasses, and social policy. *Journal of Economic Perspectives*, *25*(4), 83-94.

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.

Mayer, P. J. (1991). Inheritance of lifespan evinces no secular trend among members of six New England families born 1650–1874. *American Journal of Human Biology*, 3(1), 49-58.

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.

Mitchell, D., Hsueh, W. C., King, T. M., Pollin, T. I., Sorkin, J., Agarwala, R., Schäffer, & A. A., Shuldiner, A. R. (2001). Heritability of life span in the Old Order Amish. *American Journal of Medical Ethics, 102*(4), 346-352.

Mourits, R. J., van den Berg, N., Rodríguez-Girondo, M., Mandemakers, K., Slagboom, P. E., Beekman, M. & Janssens, A. A. P. O. (2020). Intergenerational transmission of lifespan is not affected by other familial factors: evidence from 16,905 Dutch families from Zeeland, 1812-1962. *The History of the Family, 25*(3), 484-526.

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., ... & Esko, T. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nature genetics, 54(4), 437-449.

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.

Parman, J. (2017). Gender and Intergenerational Mobility: Using Health Outcomes to Compare Intergenerational Mobility Across Gender and Over Time. Working Papers 122, Department of Economics, College of William and Mary.

Pearl, R. (1931). Studies on human lifespan. iv. the inheritance of lifespan. preliminary report. *Human Biology*, *3*(2), 245-269.

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.

Phillipe, P. (1978). Familial correlations of lifespan: An isolate-based study. *American Journal of Medical Genetics*, *2*(2), 121-129.

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

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

Preston, S. H., & Haines, M. R. (1991). 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 lifespan 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 Lifespan 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)

Tan, H. R. (2020). Did military service during World War I affect the economic status of American veterans?. *Explorations in Economic History*, *75*, 101301.

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 lifespan genetics: back to the future. *Ageing Research Reviews*, *38*, 28-39.

van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A., ... & Slagboom, P. E. (2019). Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature communications*, *10*(1), 35.

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 lifespan. *Demography*, *25*(2), 277-287.

Wainschtein, P., Jain, D., Zheng, Z., Cupples, L. A., Shadyab, A. H., McKnight, B., ... & Visscher, P. M. (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nature Genetics, 54(3), 263-273.

Ward, Z. (2023). Intergenerational Mobility in American History: Accounting for Race and Measurement Error. *American Economic Review*, 113 (12): 3213-48**.**

Wyshak, G. (1978), Fertility and lifespan in twins, sibs, and parents of twins. *Social Biology*, *25*(4), 315-330.

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)
Father's birth year	1867	1867	1867
	(14.09)	(14.05)	(14.15)
Mother's birth year	1872	1872	1872
	(13.30)	13.25	(13.35)
Birth year	1901	1901	1901
	(11.62)	(11.62)	(11.62)
White	0.99	0.99	0.99
Non-White	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

Notes: The sample includes all individuals who were age 25 or older in one of the US censuses from 1900-1920 who were successfully matched to the tree and for whom we could compute age at death. See text for further details on sample construction and sample selection.

Table 2. Lifespan quintile transition matrix, by sex

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

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 the first matrix is restricted to mothers and daughters. The sample for the bottom matrix is restricted to fathers and sons. It compares the portions of the son/father (mother/daughter) sample in a lifespan quintile given their father's/son's (mother's/daughter's) quintile N= 13,944,386 for women and 12,189,775 for men.

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)	

Notes: Each cell separately provides the estimated regression coefficient in lifespan between the two individuals indicated in the row. Errors are clustered by family. Column (1) includes no controls and regresses the child's lifespan on the parent's lifespan. Column (2) includes dummies for the parent's year of birth and for the child's year of birth. Column (3) controls for dummies indicating the state of birth of the child and the state of birth of the parent. Column 4 includes race and birth order dummies.

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

	Outcome			
Model	Lifespan (Years)	Percentile	Log Lifespan	# of obs.
Son/Father	0.090	0.090	0.076	13,944,386
	(0.0003)	(0.0003)	(0.0003)	
Son/Mother	0.062	0.078	0.048	13,944,386
	(0.0003)	(0.0003)	(0.0003)	
Son/Parents' Average	0.141	0.162	0.132	13,944,386
	(0.0004)	(0.0004)	(0.0004)	
Daughter/Father	0.075	0.079	0.059	12,189,775
	(0.0003)	(0.0003)	(0.0004)	
Daughter/Mother	0.074	0.094	0.056	12,189,775
	(0.0003)	(0.0003)	(0.0003)	
Daughter/Parents' Average	0.142	0.166	0.128	12,189,775
	(0.0004)	(0.0004)	(0.0005)	

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. Sibling Correlations

Outcome	(1) Adult longevity	(2) Education	(3) Income	(4) HH Income	(5) Adult longevity	(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 adult longevity (or of the indicated outcome) on sibling adult longevity (or on 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. The final column is restricted to people that both have a value for education (can be linked to the 1940 census) and have at least one sibling. This sample is about 13 million total. The reason these columns sum to more than that is there is overlap; sisters of sisters can also be sisters of brothers.

Table 6. Adult longevity coefficients among siblings and twins

Outcome	Siblings		Twins	
	Adult longevity sibling coefficient	IGPLF	Adult longevity sibling coefficient	IGPLF
Brother/Brother	0.134	0.084	0.183	0.078
	(0.001)	(0.001)	(0.006)	(0.004)
	4,126,499	4,680,402	31,335	62,670
Sister/Sister	0.105	0.069	0.162	0.07
	(0.001)	(0.001)	(0.007)	(0.004)
	3,102,766	3,693,559	28,020	56,040
Sister/Brother	0.035	0.077	0.05	0.062
	(0.0004)	(0.0004)	(0.005)	(0.003)
	6,988,569	8,183,995	45,628	91,256

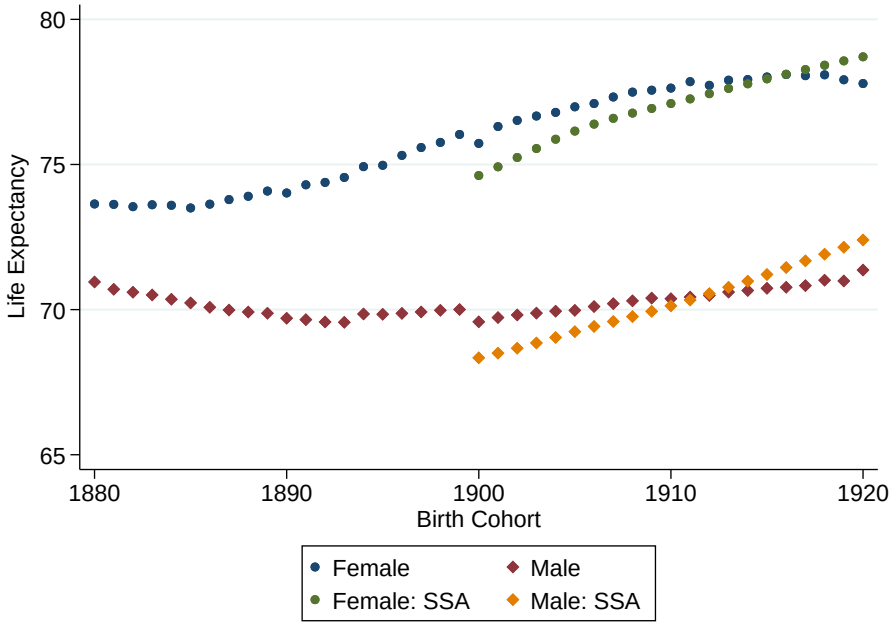
Notes: Each cell in this table is a separate regression. The sample of twins includes all pairs of individuals born in the same year and month within the same family. Columns 1 and 3 are the coefficients when sibling (or twin) adult longevity is regressed on sibling adult longevity, including birth cohort fixed effects for each person. Errors are clustered by family. The Columns 2 and 4 includes the IGPL between the children in the previous columns and their fathers.

Table 7. Variance Decompositions using Sibling Samples

	Adult longevity	Education
Panel A: Raw sibling correlations		
correlation	0.096	0.546
Panel B: Regression of adult longevity, without family FE		
R-squared	0.040	0.130
Panel C: Regression of adult longevity, 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 the 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.

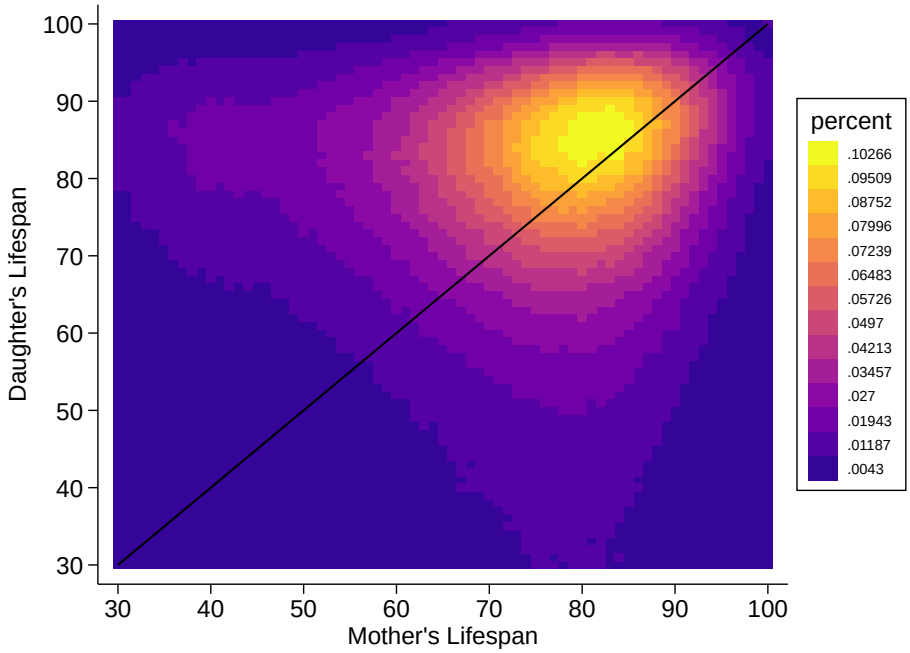
Figure 1. Trends in adult life expectancy at age 25, by sex



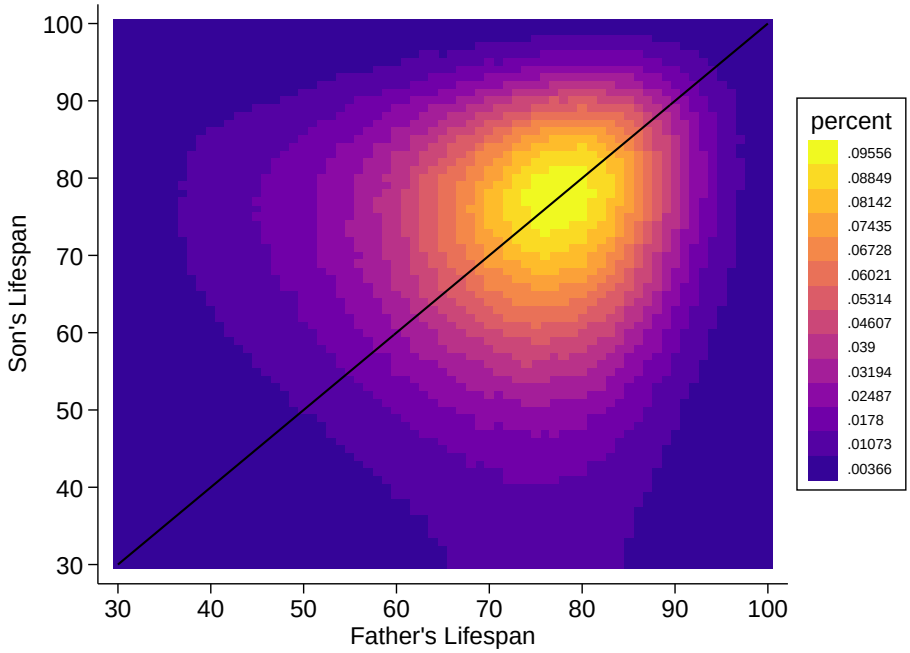
Notes: Th figure shows a cohort's adult longevity (its average/expected age at death conditional on survival to age 25) for cohorts born 1880 to 1920 who are observed in the 1900-1920 Censuses in the Census-Tree matched data and in the Social Security Administration (https://www.ssa.gov/oact/HistEst/CohLifeTables/2020/CohLifeTables2020.html)

Figure 2. Joint distribution of parent-child adult lifespan by sex, for children born 1880-1920

a. Daughter-Mother distribution of age at death



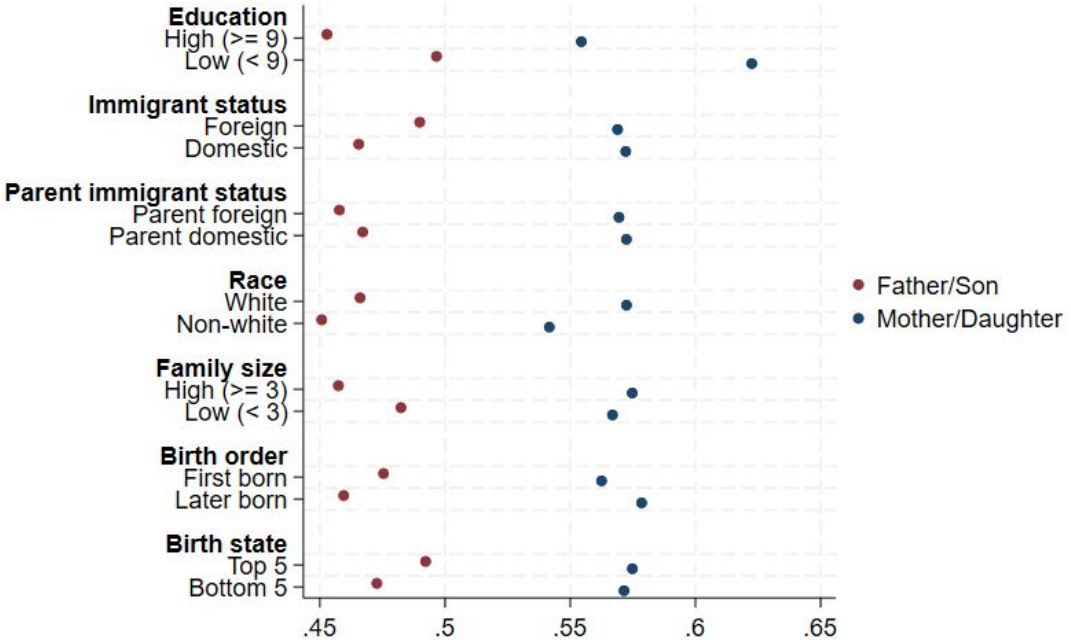
b. Son-father distribution of age at death



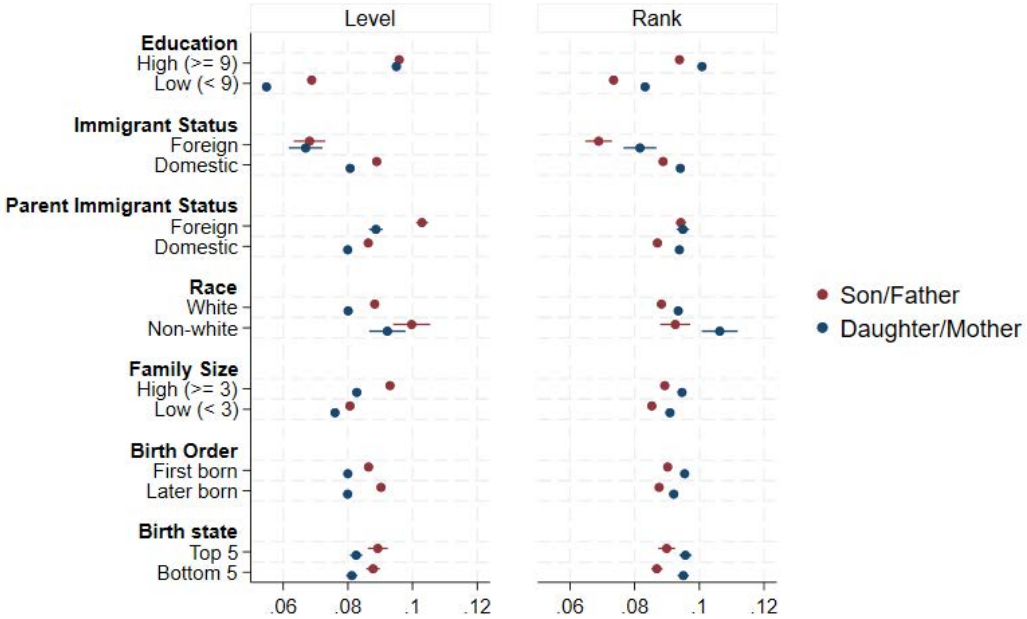
Notes: Panel a shows the distribution of the age at death among women born 1880-1920 who survived to age 25 and their mothers. Panel b shows the distribution for men and their fathers. The black line is a 45-degree line.

Figure 3. Heterogeneity in mobility in adult lifespan, by sex and group

a. Absolute mobility

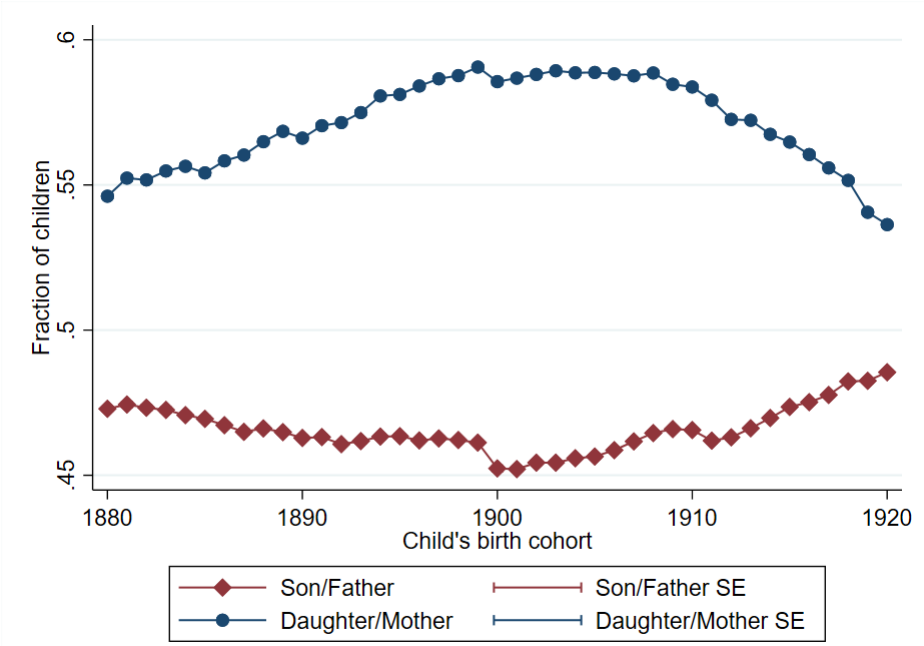


b. Relative persistence



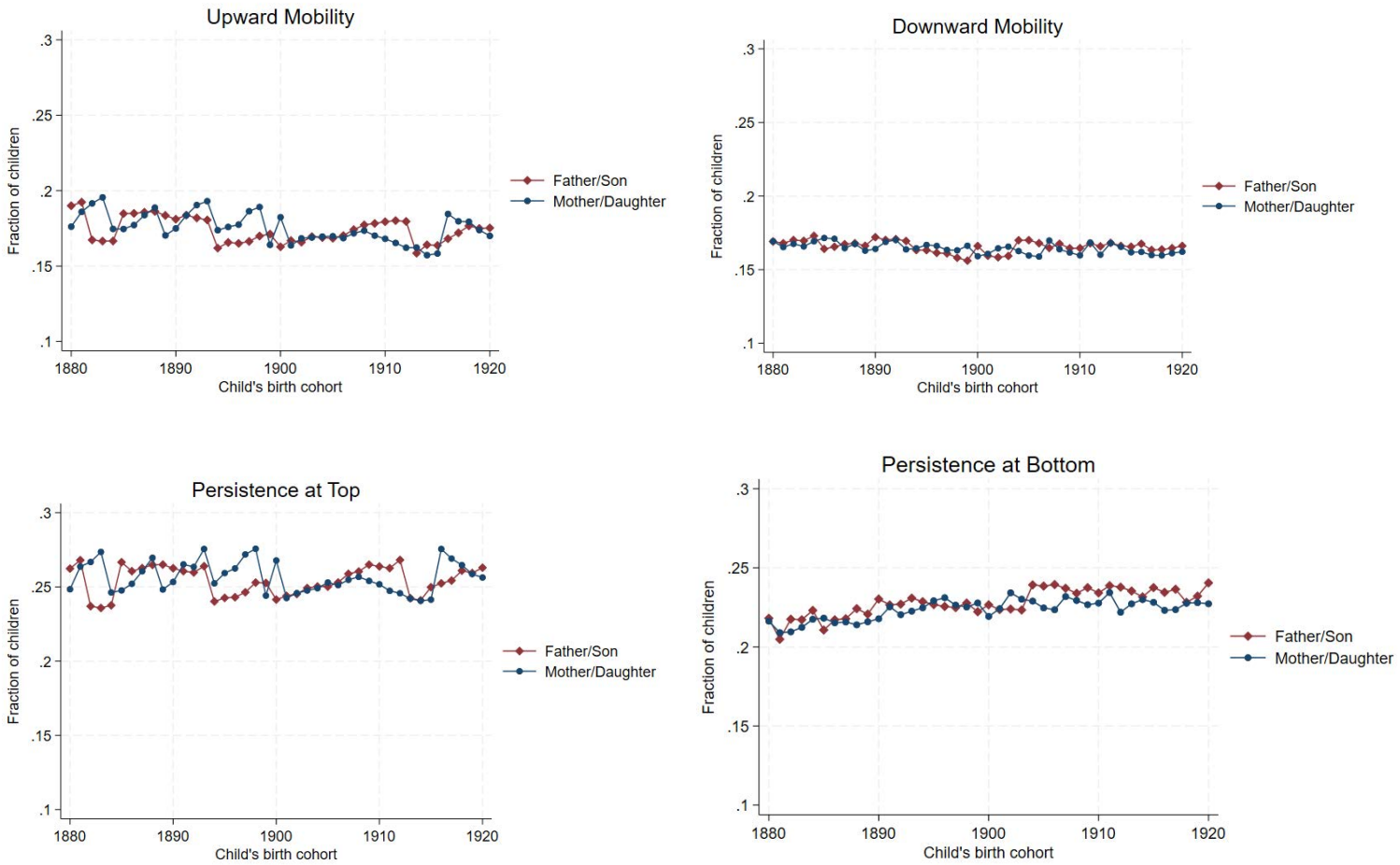
Notes: In *Panel a*, each circle reports the absolute mobility measure (the probability a child lives longer than their same-gender parent) for each of the indicated subsamples, defined by education level, immigrant status, parental immigrant status, family size, birth order or birth state. *Panel b* reports the IGPL obtained from a separate regression using only the indicated subsample, defined by education level, immigrant status, parental immigrant status, family size, birth order or birth state. Each regression includes birth cohort fixed effects for each person. Standard errors are clustered at the family level.

Figure 4. Trends in absolute and relative mobility, by sex and birth cohort
Percent of children living longer than same-sex parent



Notes: Each dot is the fraction of children in the sample whose age at death was greater than their parent's age at death (conditional on both parents and children living to age 25).

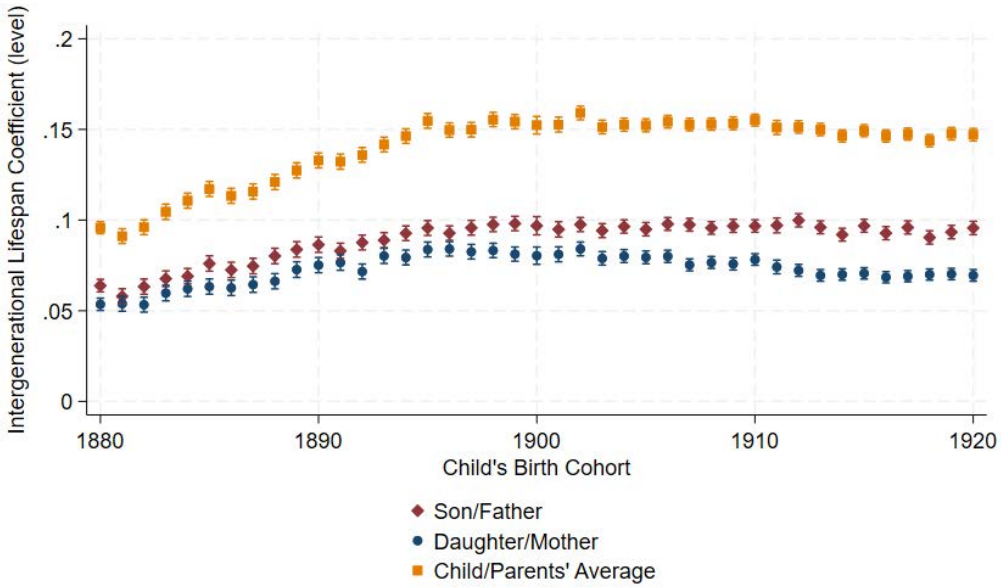
Figure 5. Trends in relative mobility in lifespan



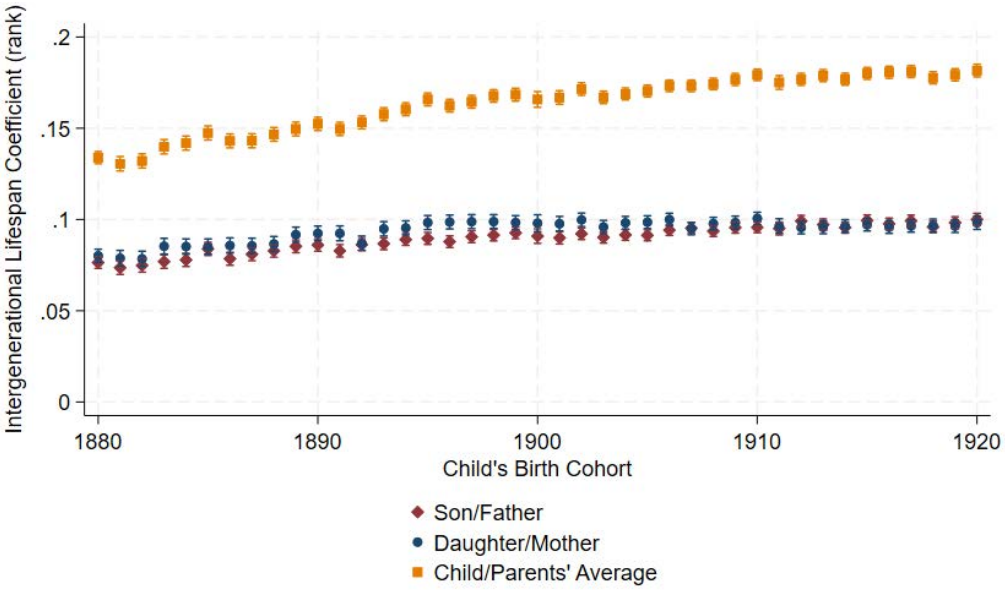
Notes: Upward mobility is the share of children whose parents where at the bottom 20 percent of the sex-specific adult longevity distribution who end up in the top 20 percent of the sex-specific adult longevity distribution in their cohort. Downward mobility is the share of children born to parents in the 20 percentile of the adult longevity distribution who end up in the bottom 20 percent of the sex-specific adult longevity distribution in their cohort. Persistence at the top is the share of children born to parents in the top 20 percent who also end up at the top 20 percent of the distribution. Persistence at the bottom is the share is children born to parents at the bottom 20 percent who also end up in the bottom 20 percent of the adult longevity distribution of their cohort.

Figure 6. Changes in the IGPL Over Time by Level and Rank
Census-Tree data for children born 1880-1920

a. Level

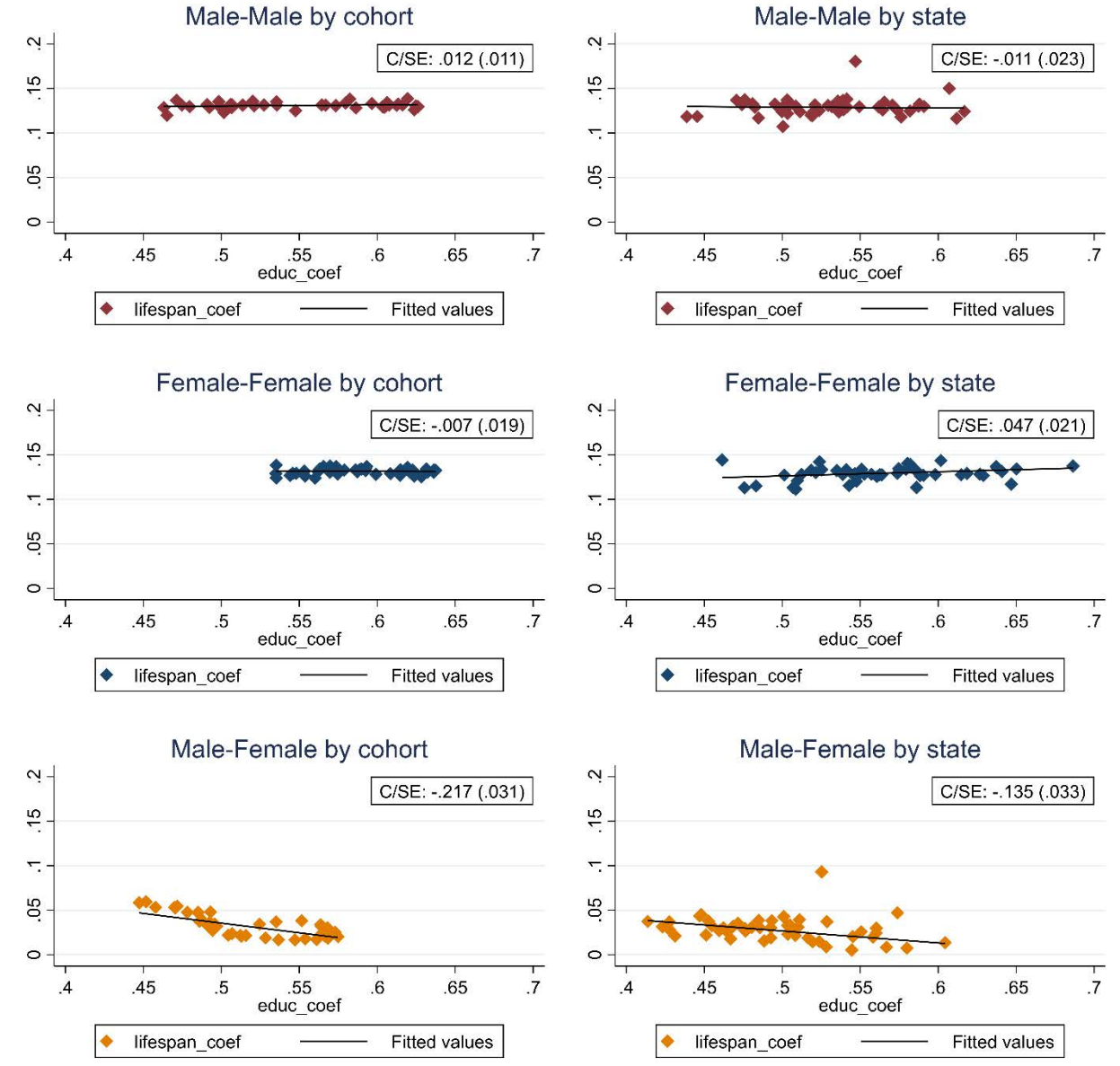


b. Rank



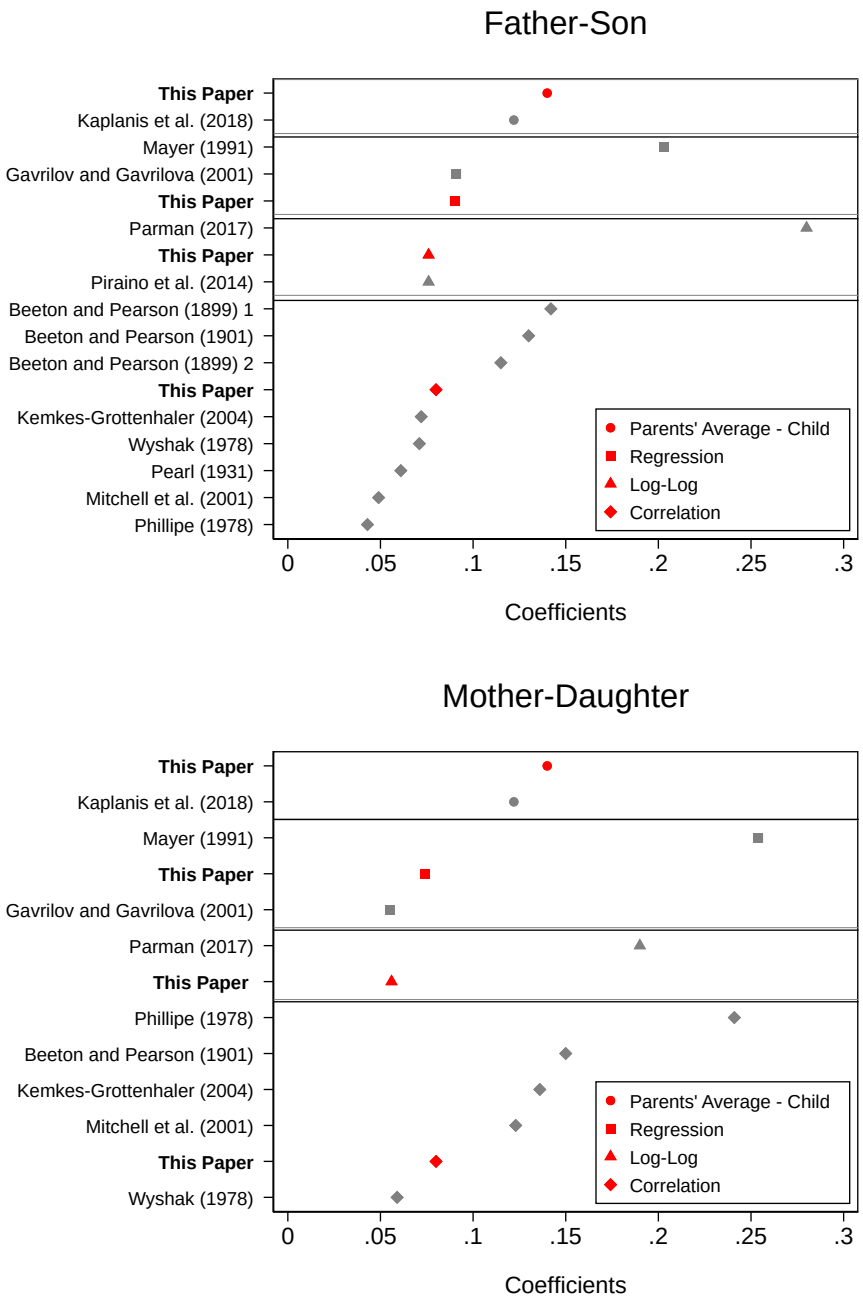
Notes: In this figure, for each birth cohort, we estimate the intergenerational coefficient of lifespan for son/father pairs, daughter/mother pairs, and the correlation between the child's lifespan and the average lifespan of both parents. Each point corresponds to the regression coefficient of a regression of the child's lifespan on a measure of the parent'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 7. Correlations Across Cohorts and States in the Sibling Associations in Adult Lifespan and the Sibling Associations in Education



Notes: These figures plot the sibling regression coefficients in longevity on the y-axis against the sibling coefficient 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 60.

Figure 8. Comparison to Previous Estimates of IGPL



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 any age restriction.

- Father-Son (Mother-Daughter) correlation for Gavrilov and Gavrilova (2001) is from Table 5 (Table 6), which uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 uses 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 1: A Simple Model of the Determination of Lifespan

Below, we provide a simple model to provide a framework for better understanding the intergenerational correlations in lifespan and their relationship to sibling correlations. Previous research on the determinants of lifespan suggests that lifespan 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 lifespan.[1]

We denote L_{ijsc}^{g} 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_{ijsc}^{g}), 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_{jsc-1}^{f} and G_{jsc-1}^{m} denote the genes of the female parent (mother) and the male parent (father).[2] Each child in family *j* receive these genes along with variation in the genes they draw from each parent (η_{ijc}^{f} and η_{ijc}^{m}).[3] Children receive half of their genes from each parent: $G_{ijsc}^{g} = \frac{1}{2}\left[G_{jsc-1}^{f} + \eta_{ijc}^{f} + G_{jsc-1}^{m} + \eta_{ijc}^{m}\right]$. We assume G_{jsc-1}^{f} and G_{jsc-1}^{m} are random variables with an unknown variance-covariance matrix reflecting the extent of assortative mating that is gene-based. The variables η_{ijc}^{f} and η_{ijc}^{m}, 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:

$$L_{ijsc}^{g} = \delta\frac{1}{2}\left[G_{jsc-1}^{f} + \eta_{ijc}^{f} + G_{jsc-1}^{m} + \eta_{ijc}^{m}\right] + \alpha_{sc} + \gamma_i^{g} + \theta_j^{SES} + e_i \qquad (1)$$

where e_i is a random variable, assumed to be independent across individuals, and from *G*, α_{sc},

[1] Throughout, we use the terms lifespan, longevity 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 the duration of life which is referred to as lifespan (van den Berg et al. 2017).
[2] Gender captures biological, hormonal and social effects.
[3] Here *c-1* represents the parents' cohort.

γ_i^g, and θ_j^{SES}. The parameter δ represents the effect of genes on lifespan, where the genes are assumed to affect lifespan through a linear index, such as a polygenic score.[4]

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.[5] 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.[6] We are also ruling out interactions between gender and the environment.[7] We discuss the implications of these assumptions later, and we relax them in estimation. But this simple model provides a useful baseline.[8] This data-generating model has strong empirical implications, summarized here in proposition 1 (the Appendix outlines the covariances in lifespan 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

[4] In this framework *G*, or more precisely the function $\delta\frac{1}{2}\left[G_{jsc-1}^f + \eta_{ijc}^f + G_{jsc-1}^m + \eta_{ijc}^m\right]$, can be viewed as a polygenic score that combines all the genes that are known to determine lifespan. 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 lifespan. Timmers et al. construct a polygenic score to predict lifespan using the identified genes.

[5] 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.

[6] 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).

[7] 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).

[8] This model has some important limitations. The model is mechanistic. The only reasons 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 lifespan or to optimally choose the locations that would maximize their lifespan. We also exclude the possibility that the lifespan of the parents has a direct impact on the lifespan 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 lifespan of individuals (which empirical evidence shows matters) – we assume only the family's SES matters.

controlling for cohort effects in the basic regression.)

b. *The simple bivariate intergenerational correlation in lifespan 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 lifespan, and the variance of the gender component. Thus, this correlation does not uniquely describe the extent to which genes affect lifespan, 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* $V(G_{jsc-2}^{paternal\ m})$.[9]

[9] 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.

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

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

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.[10] Similarly, it is also unclear whether brother covariances will exceed sister covariances.[11]

The closed population assumption assumes there is no significant in- or out-migration of genetically diverse individuals.[12] 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.[13]

[10] 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).

[11] 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.

[12] 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.

[13] 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.

Appendix Tables and Figures

Table A.1. 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		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

	Mother-Daughter: 0.106-0.241				
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 (2020)	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 A.2. 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

Notes: The table shows the remaining years of life left at different ages. 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.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/

Table A.3. Comparison of Matched Sample and Full Census Sample Mean Values

Variable	Matched Analysis Data	Unmatched Census data
Lifespan	73.0	
Female	0.47	0.52
Birth Year	1901	1898
White	0.99	0.82
Black	0.01	0.14
Northeast	0.15	0.24
Midwest	0.41	0.25
South	0.35	0.35
West	0.07	0.04
Immigrant	0.01	0.08
Father is Immigrant	0.13	0.24
Mother is Immigrant	0.10	0.19
Observations	26,134,161	960,504,392

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.

Table A.4. Raw Correlations in Lifespan Across Generations

	Outcome			
Model	Lifespan (Years)	Percentile	Log Lifespan	Observations
Son/Father	0.08	0.09	0.06	13,944,386
Son/Mother	0.06	0.08	0.05	13,944,386
Son/Parents' Average	0.10	0.12	0.08	13,944,386
Daughter/Father	0.06	0.08	0.05	12,189,775
Daughter/Mother	0.08	0.09	0.06	12,189,775
Daughter/Parents' Average	0.10	0.12	0.08	12,189,775
Father/Mother	0.05	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 A.5. Summary Statistics of Sibling Subsample

	Census based sample matched to FamilySearch, cohorts born 1880-1920	
	(1) Full Sample	(4) 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 A.6. Sibling correlations for outcomes in the 1940 census

Model	(1) Adult longevity	(2) Education	(3) Income	(4) HH Income	(5) Adult longevity	(6) Education	(7) 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 A.7. Intergenerational Persistence of Education and Lifespan

	Father		Mother	
	Education	Lifespan	Education	Lifespan
Son	0.440	0.175	0.497	0.140
	(0.000)	(0.001)	(0.001)	(0.001)
	3,184,950	3,184,950	3,442,355	3,442,355
Daughter	0.402	0.126	0.483	0.133
	(0.000)	(0.001)	(0.001)	(0.001)
	2,615,551	2,615,551	2,796,741	2,796,741

Notes: Sample is restricted to individuals who survive to age 25. We also restrict the sample to have both education and lifespan for comparability.

Table A.8. 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 A.9. 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

Notes: This table was created using a previous iteration of the dataset. The left two columns are copied from Table 6. The right two columns have the same specification as the left but are restricted to a subsample that also matched to a death certificate record on Family Search.

Figure A.0. Data Construction

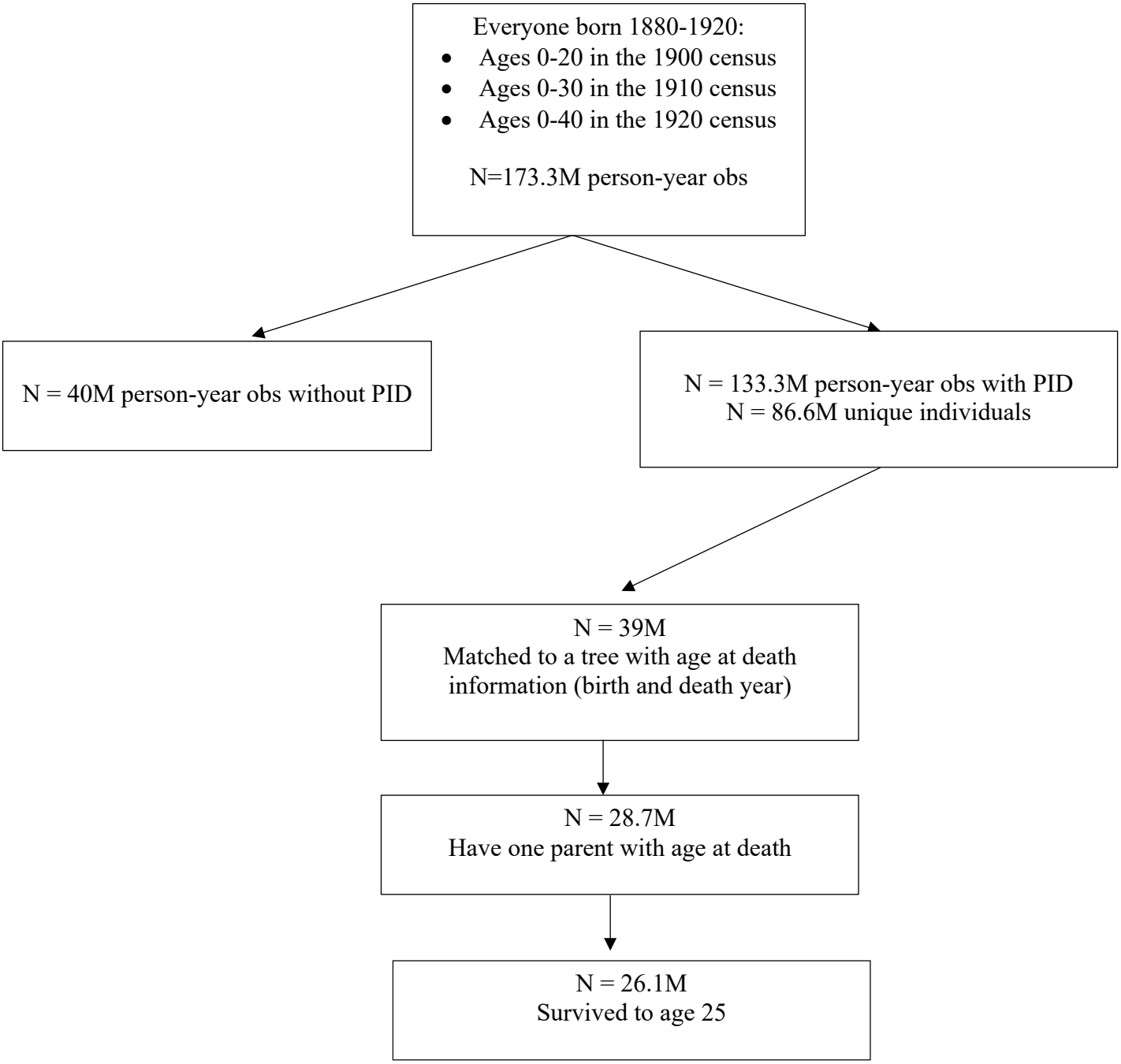
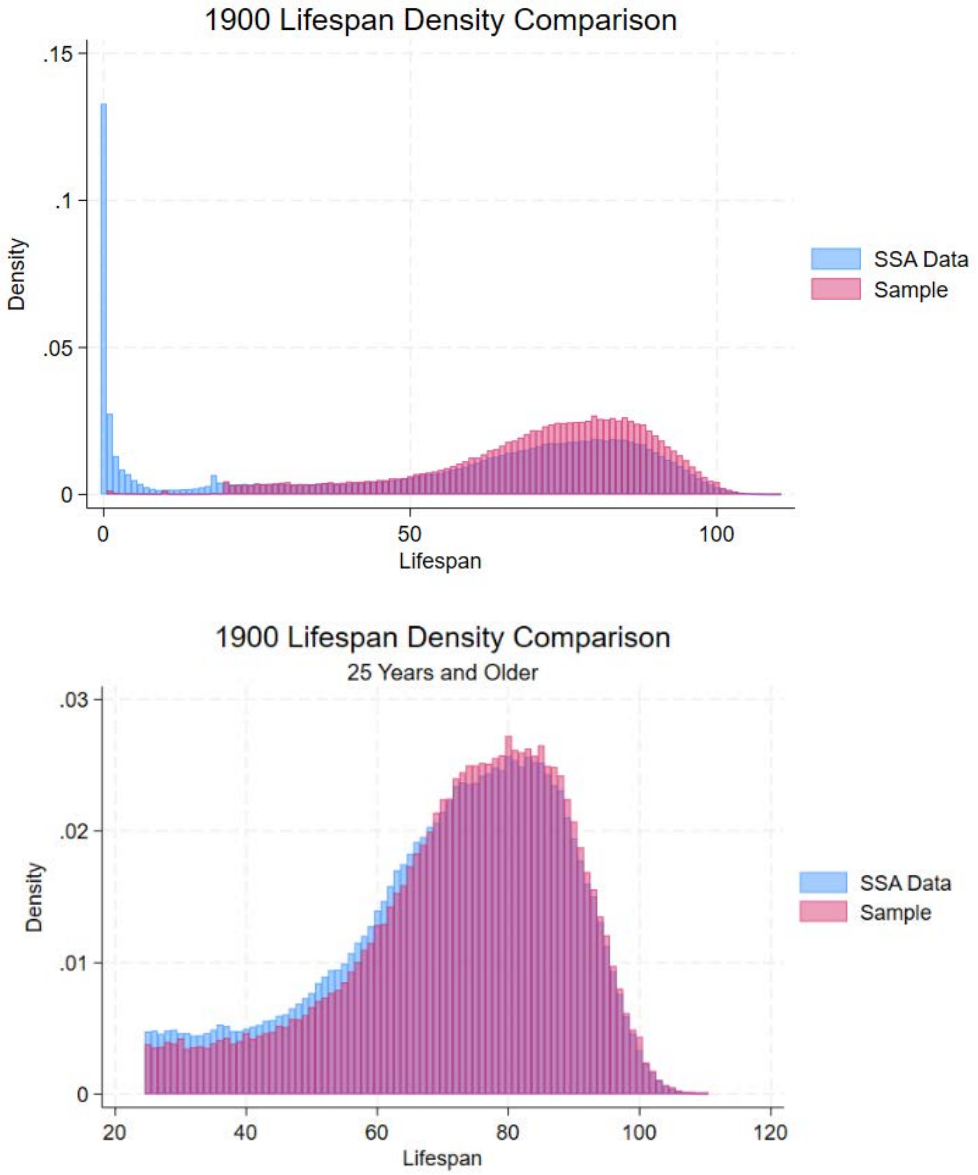
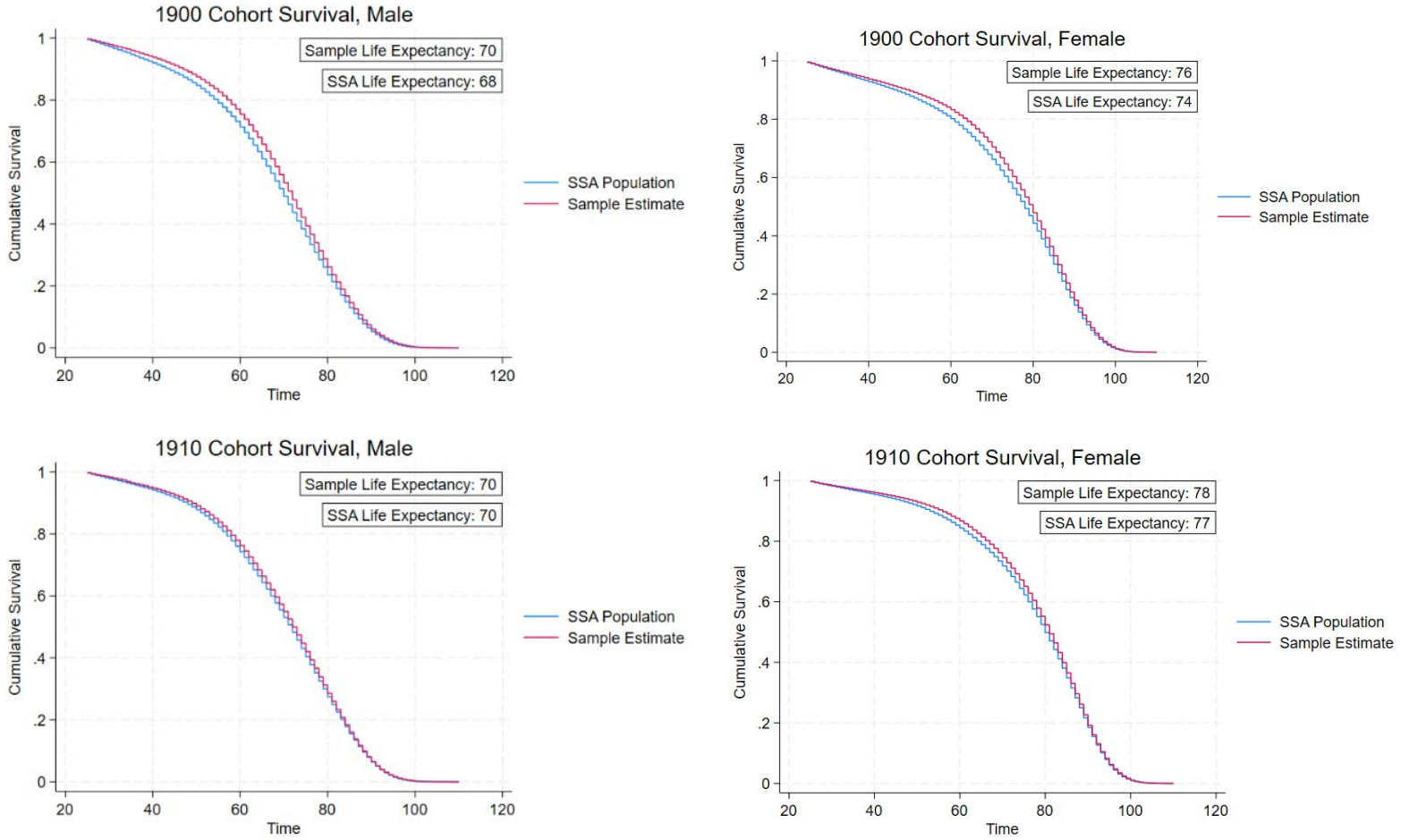


Figure A.1. Distribution of the age at death In the Census-Tree data
Comparison with SSA Cohort Data



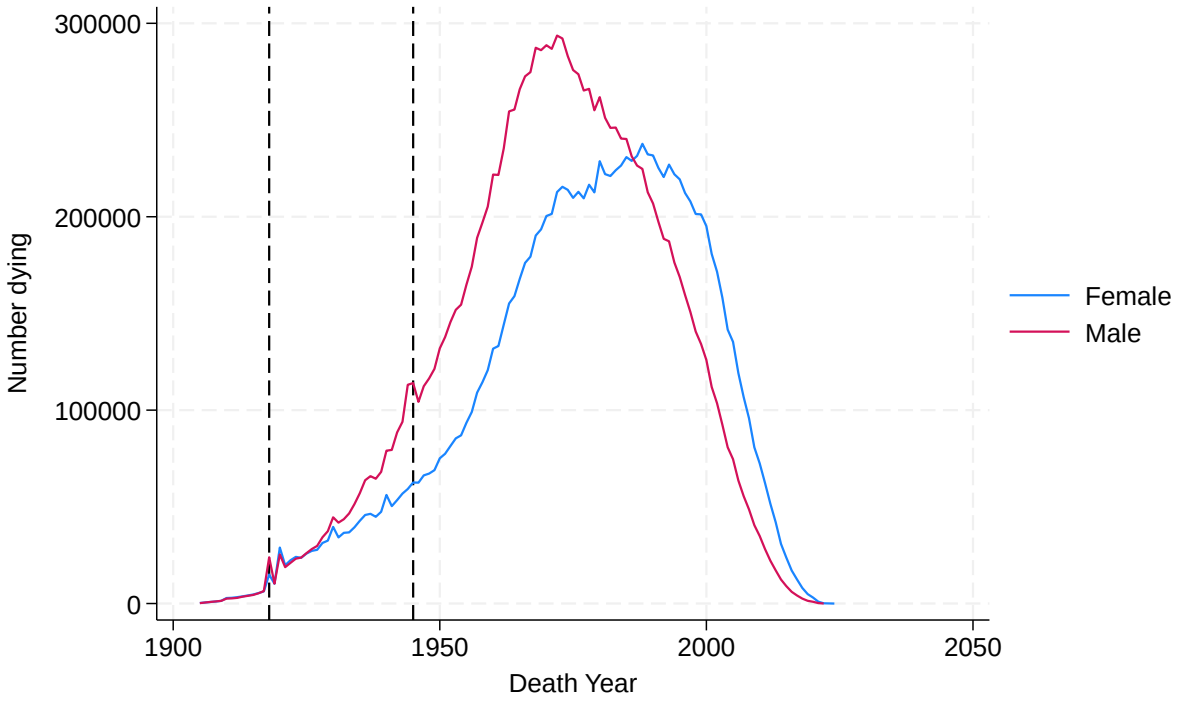
Notes: These figures use our sample derived from the Family Tree (see text for details) and cohort life tables produced by the Social Security Administration (SSA), available here: https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf.

Figure A.2. Comparing survival rates in the Census-Tree data and the SSA cohort data



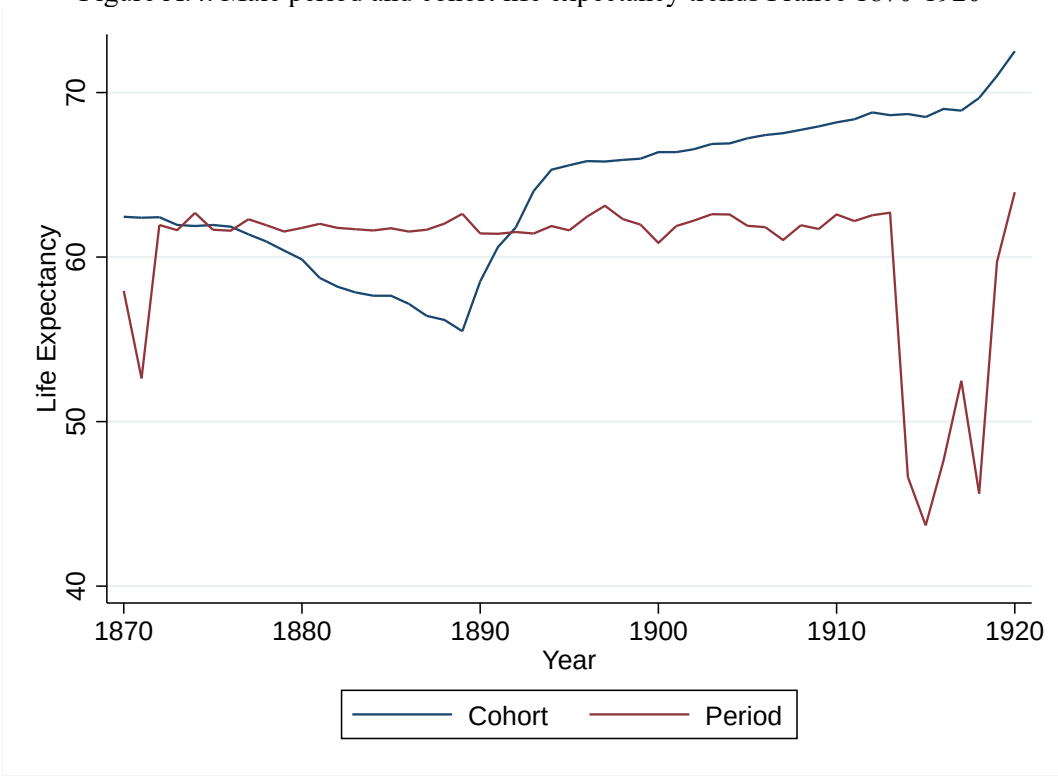
Notes: These figures use our sample derived from the Family Tree (see text for details) and cohort life tables produced by the Social Security Administration (SSA), available here: https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf. Kaplan-Meier estimates are produced using the methods described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

Figure A.3. Number of deaths by death year in our sample.
Cohorts born 1880-1920 surviving to age 25.



Notes: The figure shows the death year of the individuals born 1880-1920 who survived to age 25 and who are in our sample (they have birth and death dates and so do their parents). The dahsed lines denote 1918 and 1945, the deadliest years of WWI and WWII. In 1918 there was also a flu pandemic.

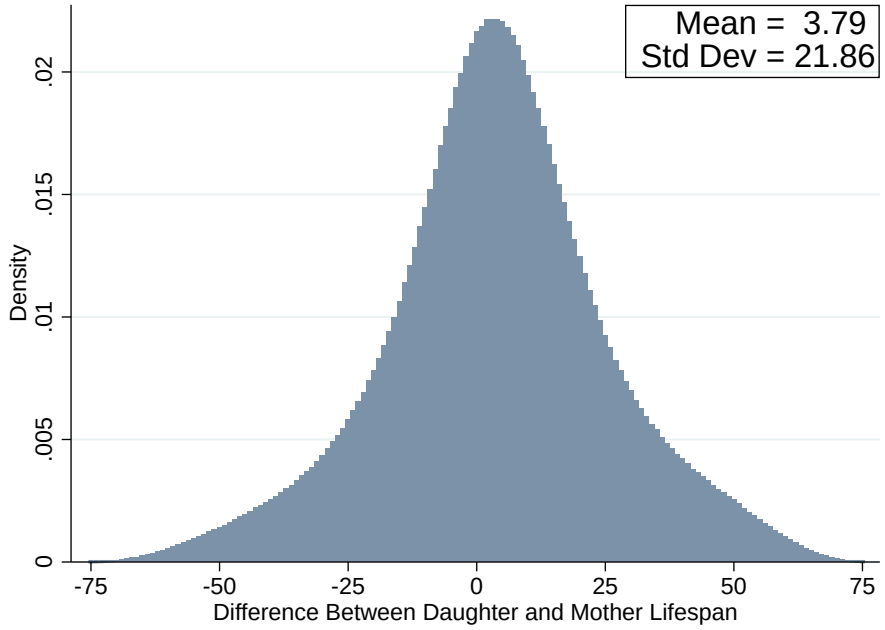
Figure A.4. Male period and cohort life expectancy trends France 1870-1920



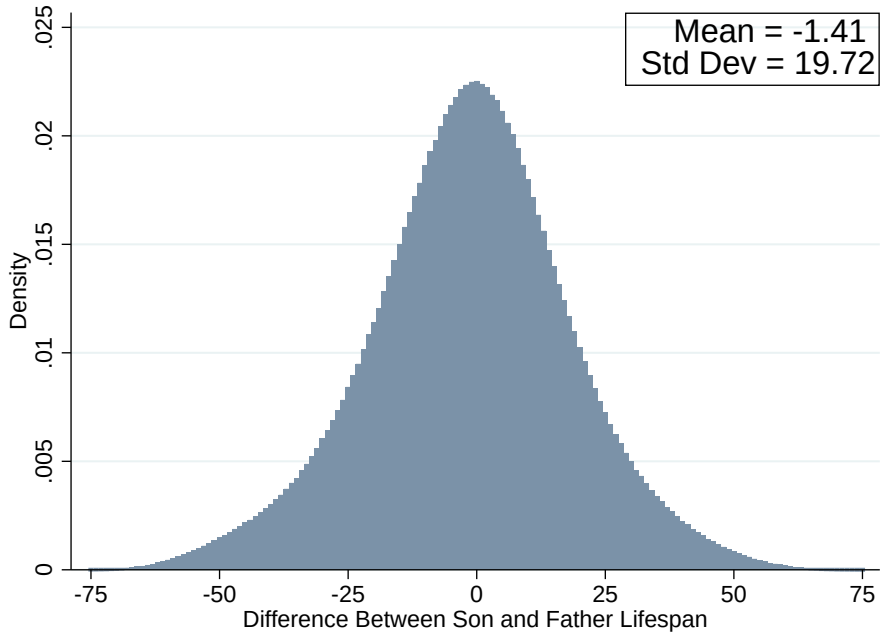
Notes: Uses period and cohort tables from the Human Mortality Database.

Figure A.5. distribution of adult lifespan gap in Census-tree data

a. Daughter age at death – mother age at death



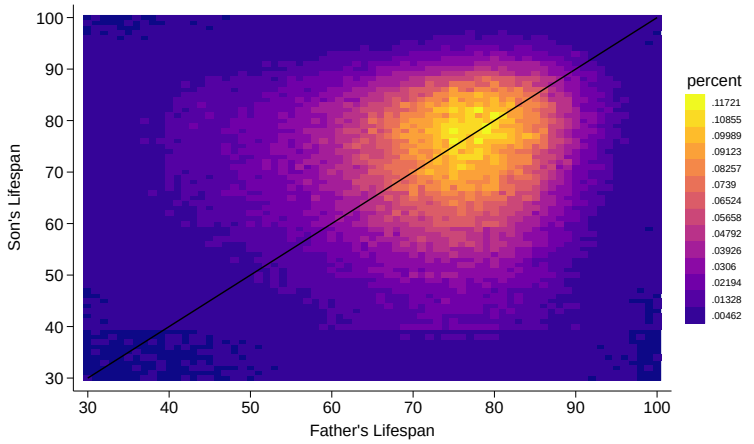
b. Son age at death – father age at death



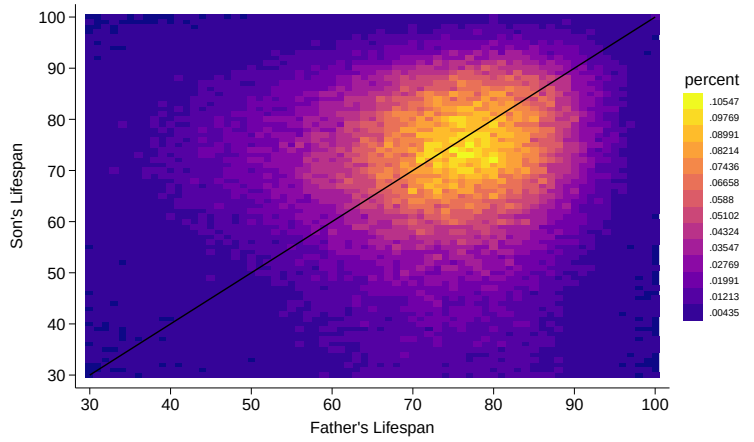
Notes: Figure shows the gap in the age at death of the child minus the age at death of the parent, conditional on both parent and child surviving to age 25. Census-Tree data for cohorts born 1880-1920.

Figure A.6. Heat maps by cohort for men

a. 1880



b. 1900



c. 1920

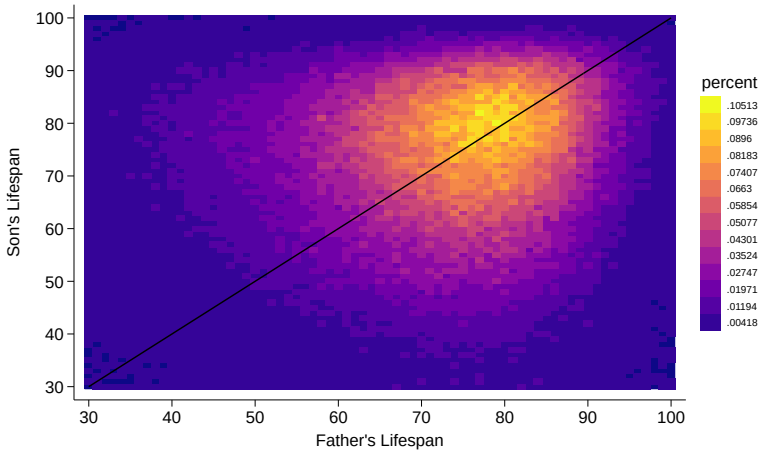
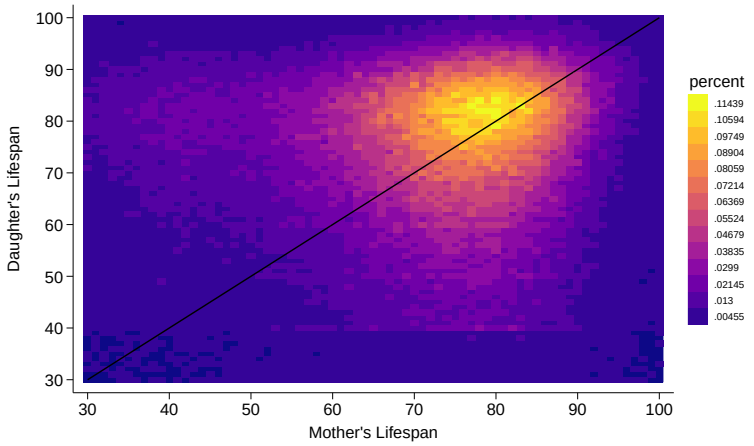
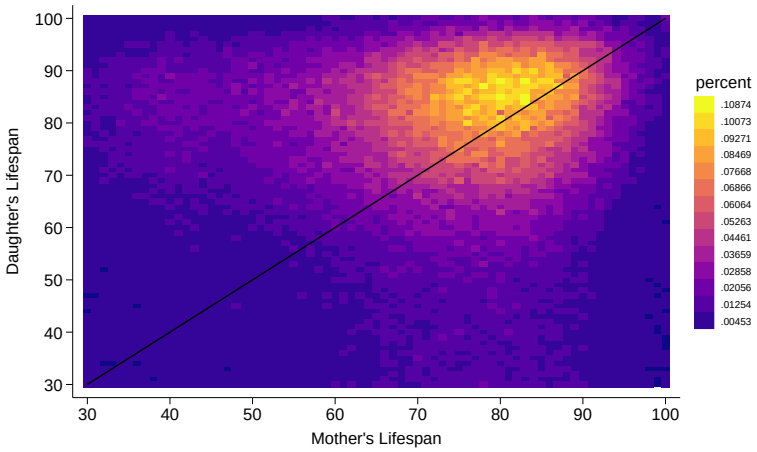


Figure A.7. Heat maps by cohort for women

a. 1880



b.1900



b. 1920

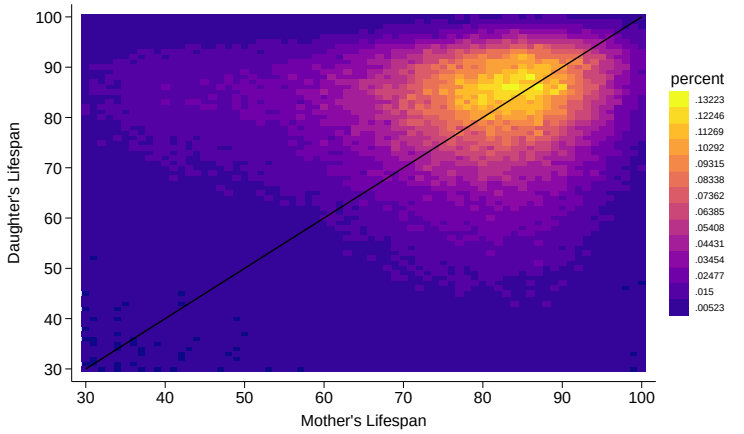
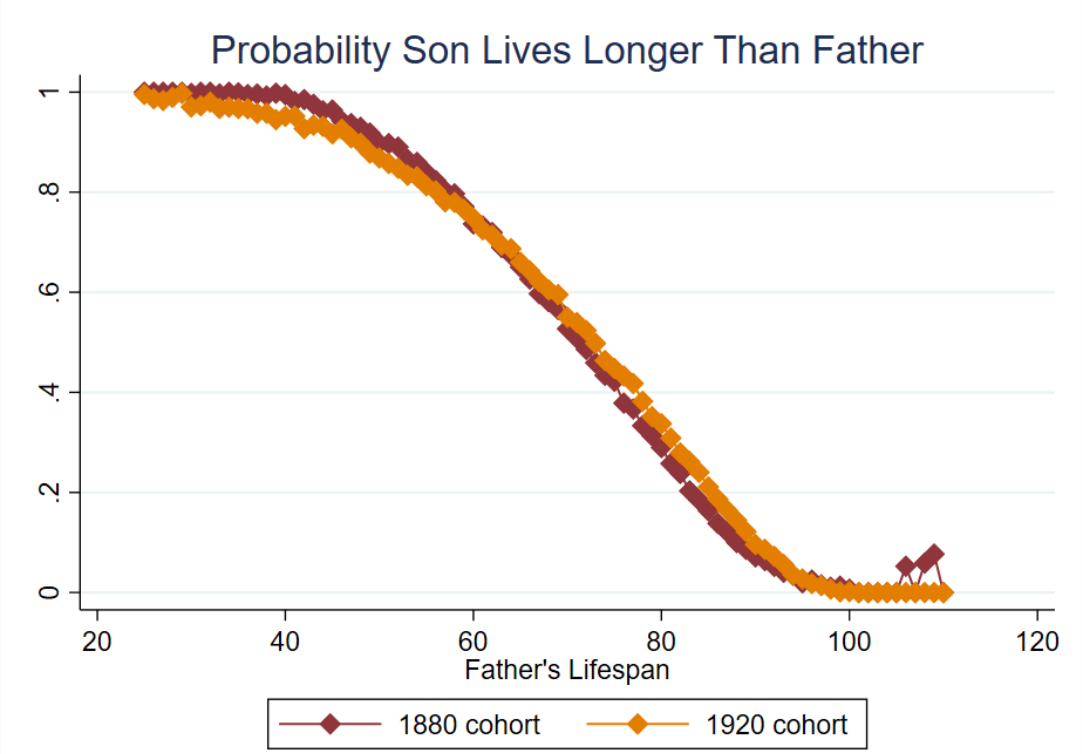


Figure A.8. Changes in Absolute Mobility 1880-1920



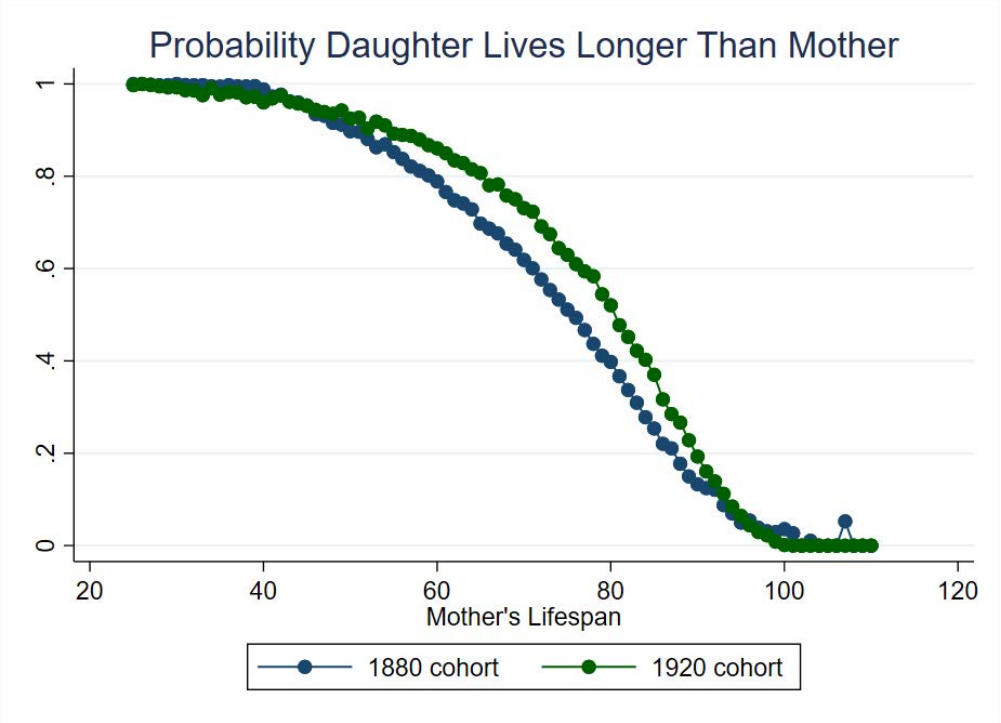
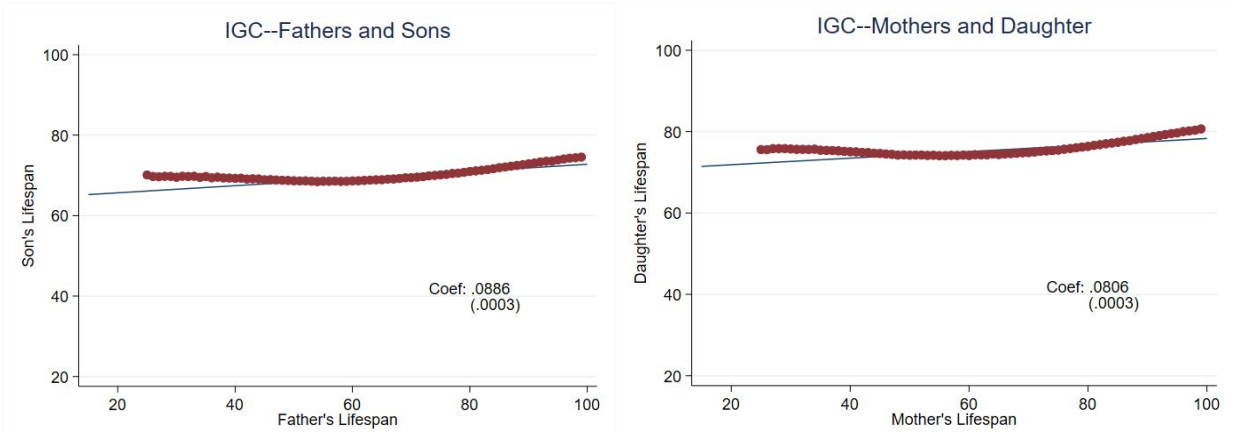
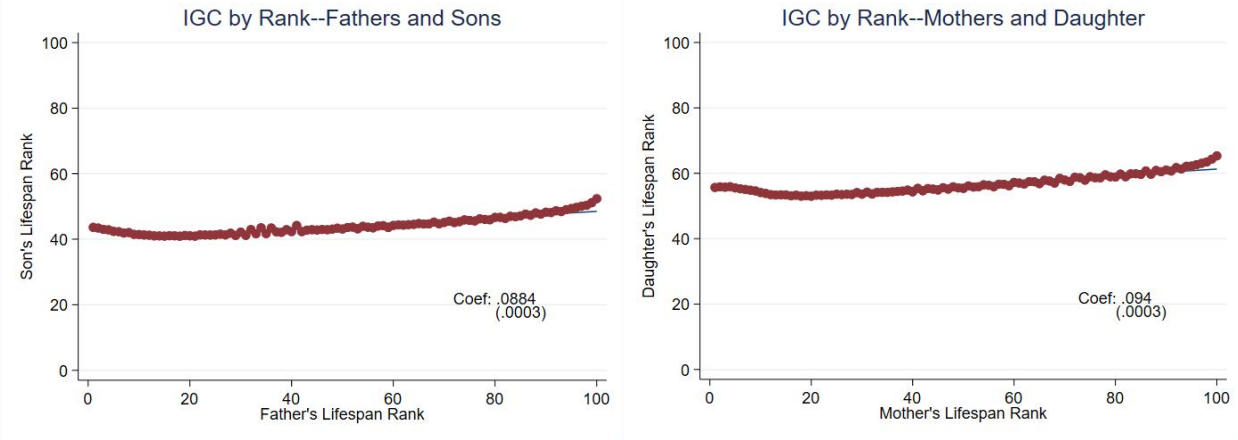


Figure A.9. Test for Linearity of the child-parent lifespan relationship, by sex

a. IGPL in levels: Child adult lifespan as a function of parent lifespan

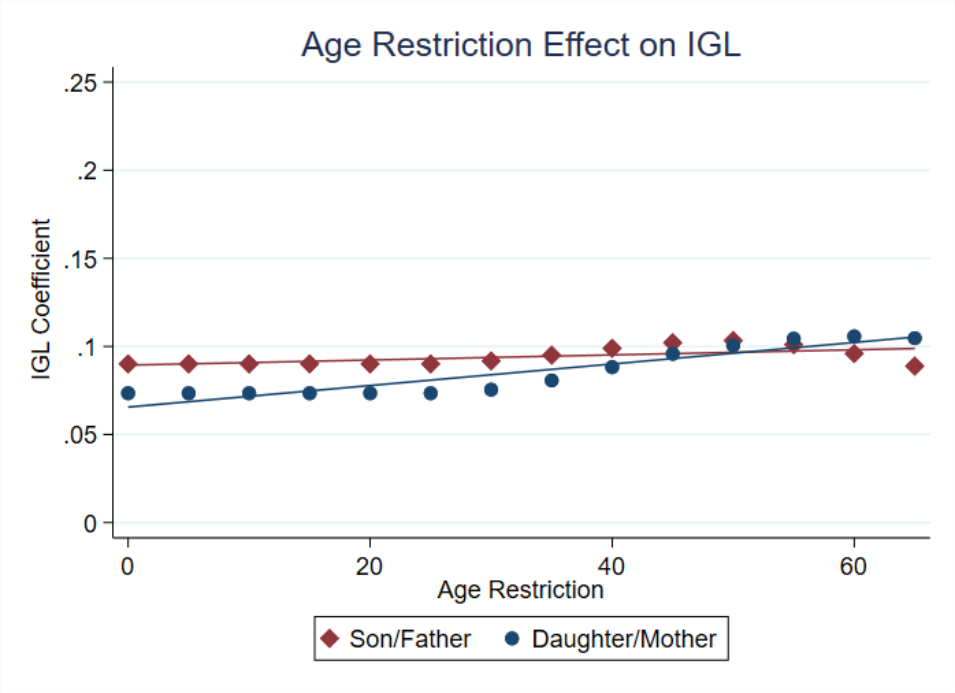


b. IGPL in ranks: Child adult lifespan percentile as a function of parent adult lifespan percentile



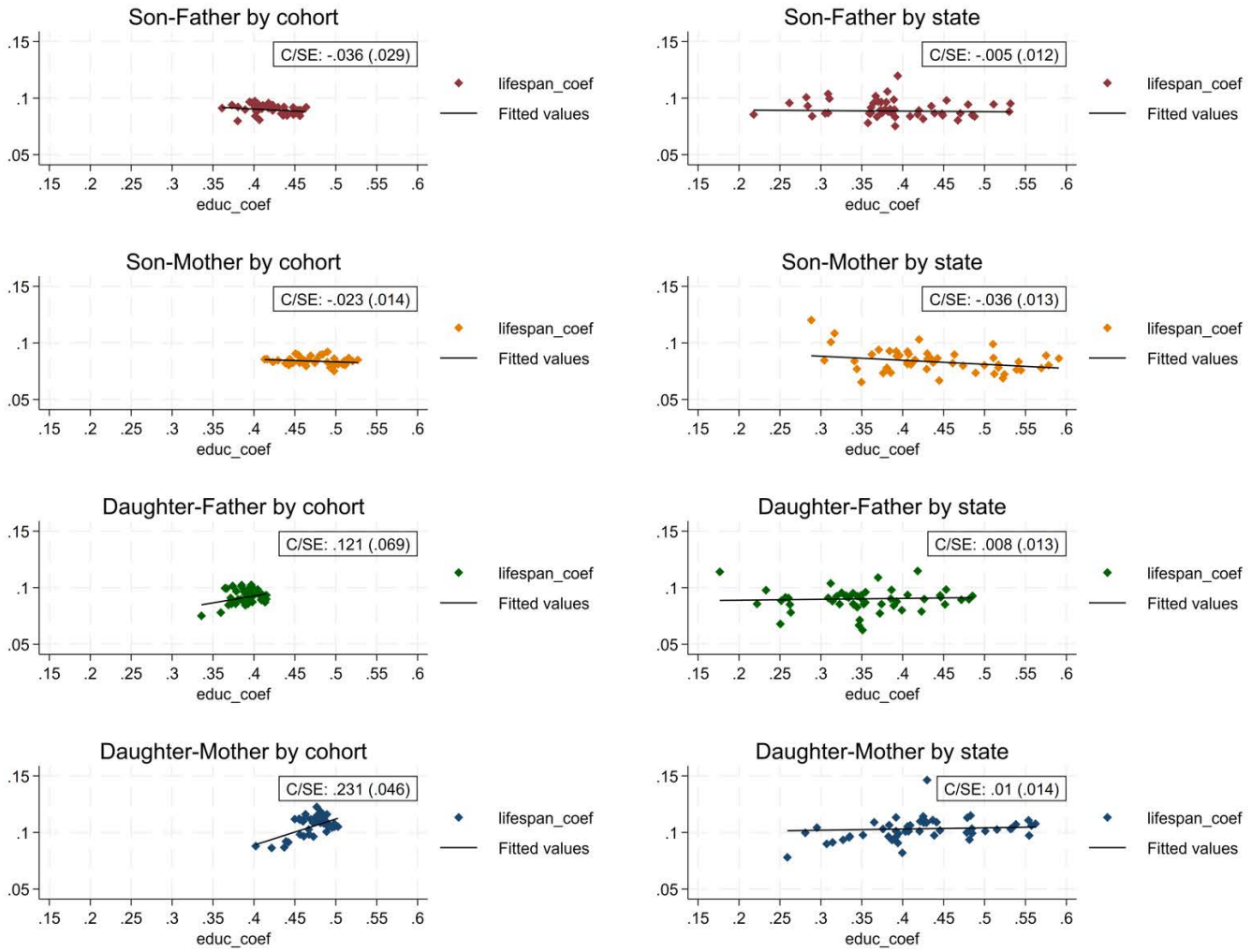
Notes: The top figures plot the average of the son's lifespan in one-year bins based on the father's lifespan, conditional on both parents and children living to age 25. The bottom figures relate the average son's (daughter's) percentile in the distribution of the age at death among sons, relative to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

Figure A.10. 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 both parent and child.

Figure A.11. Correlations Across Cohorts and States in the Persistence of Adult Lifespan and the Persistence of Education



Notes: These figures plot the regression coefficients of child longevity on parent longevity on the y-axis, against the regression coefficient of child education on parent 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 60.