

Intergenerational Transmission of Lifespan in the US*

Sandra E. Black
University of Texas at Austin and NBER

Neil Duzett
Texas A&M

Adriana Lleras-Muney
UCLA and NBER

Nolan Pope
University of Maryland and NBER

Joseph Price
BYU and NBER

March 2026

Abstract: We examine the transmission of lifespan across generations in the US using a unique dataset containing more than 26 million individuals born between 1880 and 1920. On average, 47 percent of men and 57 percent of women lived longer than their parents, though this varied across cohorts and socio-economic groups. The intergenerational persistence in lifespan is low across cohorts and socio-economic groups, and it is much smaller than persistence in socio-economic status. Moreover, persistence in lifespan and in socio-economic status are independent of each other. Lifetime well-being, which combines socioeconomic and lifespan measures, is less persistent than socio-economic measures suggest.

* 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, Columbia University, UCLA's Demography conference, London Business School's Economics of Longevity and Aging conference, the University of Maryland, the University of Melbourne, CUNY, Harvard Kennedy School, NYU Stern, University of Melbourne, Hunger College, the Federal Reserve Bank of New York, USC, the Harris School of Public Policy, NHH, and the University of Wisconsin for helpful comments. All errors are our own.

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 (Black and Devereux, 2011; Stuhler 2018). However, if we ultimately care about intergenerational persistence in overall well-being, looking at socioeconomic status alone may be insufficient. Lifespan is a key component of lifetime well-being: all else equal, lifetime well-being is greater when individuals live longer. Previous research has found that individuals highly value increases in lifespan (Jones and Klenow, 2016; Becker, Philipson, and Soares 2005). Given that, ultimately, it is the intergenerational transmission in well-being that we care about, understanding the intergenerational persistence in lifespan, and how it relates to the economic components of well-being, is fundamental. While we know that there is considerable and rising cross-sectional inequality in lifespan – with richer, more educated individuals living substantially longer lives in the US and most other developed countries (e.g. Cutler et al. 2011, Marmot 2015) – it is less well-known how much these differences are transmitted from parents to children, despite the fact that lifespan is an essential determinant of well-being.

We use a newly compiled dataset to study the transmission of lifespan across generations, its relation to socio-economic status (SES), and its effect on the persistence of well-being. We combine United States Census records with data from the wiki-style Family Tree created by FamilySearch, which includes over 1.4 billion people and is the largest collection of its kind. We start with all individuals observed in the 1900, 1910, and 1920 decennial 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 family members, focusing on individuals who survived to at least age 25. Our final sample includes 26.1 million individuals matched to their parents and siblings. This is the largest sample ever constructed to study the intergenerational transmission of lifespan; it includes more than 30% of the US population born between 1880 and 1920 who survived to age 25.

We use these data to document a new series of facts about lifespan transmission in the US. We first develop a simple model to provide insight into the determinants and persistence of lifespan across generations. Then, we examine both absolute and relative mobility to provide a comprehensive picture of intergenerational mobility in lifespan across cohorts and for different sub-populations. Absolute mobility refers to whether children live longer than their parents,

whereas relative mobility refers to whether parents who live longer than the average in their cohort will have children who live longer than the average in their cohort. During the late 19th century and the first half of the 20th century, incomes, educational attainment, and life expectancy were all 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 lifespan mobility was similarly low and whether it followed the same trends.

Then, to better understand how lifespan mobility and economic mobility are related, we conduct two analyses. First, we compare the transmission of occupation-based income to that of lifespan for men born from 1880 to 1920. Second, we look at correlations between siblings by constructing a subsample of siblings alive in 1940 for whom we can observe education, income, and occupation in addition to lifespan. In our simple model, if the environment and SES siblings share are more similar than the environment and SES that parents and children share, then sibling-sibling correlations serve as an upper bound on the parent-child correlations. We also investigate how these parent-child and sibling-sibling correlations, in lifespan and SES, are related across cohorts and space. We end by combining lifespan and SES measures into a lifetime utility measure to study the intergenerational transmission of well-being.

We find that mobility in lifespan is high, although different measures – particularly when comparing absolute versus relative mobility – result in different patterns across sex and across cohorts, with relative mobility exhibiting much less variation than absolute mobility measures. Our persistence estimates are consistently low across different subgroups, including sex, race/ethnicity, immigration status, place of birth, education, and family types. In contrast, we find that persistence in SES is high in our data. Intergenerational persistence in occupation-based income percentiles hovers around ~ 0.35 , compared to persistence of ~ 0.1 for lifespan. Among siblings, 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 suggest that lifespan and SES measures paint different pictures of persistence, *even within the same sample*. Strikingly, we find that intergenerational persistence in occupation-based income and intergenerational persistence in lifespan are independent of each other in space and across cohorts. The same is true of sibling correlations in education and sibling correlations in lifespan. In other

words, the times and places where parents have transmitted their SES to their children are not the same as those where they have transmitted their lifespan.

We then compute lifetime utility, using SES measures and adult lifespan and focusing on men for whom SES is well defined in historical records. Our results show that measures of the intergenerational transmission in well-being that incorporate both SES and lifespan show much more mobility than measures of SES alone, at least for men born 1880-1910. They also show larger increases across cohorts than either lifespan or socio-economic correlations, suggesting economic measures of mobility are not sufficient for characterizing mobility in well-being.

We end by investigating why the intergenerational persistence in lifespan is so low, given that there are important genetic, behavioral and social determinants of health (and thus lifespan) that previous research has shown to be strongly transmitted from parents to children. To do so, we examine twins, among whom the genetic component is likely to be large. Even among twins, however, correlations in lifespan are low. 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 several literatures in economics and demography on the intergenerational persistence in health, broadly defined. There is a burgeoning literature in economics examining the intergenerational persistence in health which generally examines specific measures of health, such as birthweight (Currie and Moretti, 2007; Giuntella et al. 2022), infant mortality (Lu and Vogl, 2023), mental health (Johnston et al. 2013), BMI (Classen 2010; Classen and Thompson 2016), or chronic conditions (Thompson 2014) such as anemia (Kumar and Nahlen 2023). 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 (e.g. Andersen, 2021; Halliday et al. 2019, 2021) have focused on constructing more comprehensive health indices that can be compared across generations and that summarize all aspects of health at a meaningful point in the lifetime (or across the lifetime) – see the recent handbook chapter by Halliday (2023) for an excellent summary of this work. Our work focuses on lifespan, a complementary measure of health that focuses on the quantity, and not quality, of life. 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. It is also of interest itself, since the length of life directly affects lifetime utility. We further contribute to this literature by examining 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 mobility 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 (Beeton and Pearson, 1899, 1901). A complete list of these studies along with a summary of their findings can be found in Table A.1. Because this literature focuses primarily on genetic influences, it typically uses small convenience samples, making results hard to generalize, and comparisons across cohorts and groups difficult. Our samples are an order of magnitude larger than those used in this literature and include information that can be used to assess its representativeness.¹ Nevertheless, our results are similar, which we show can be attributed to low levels of heterogeneity across all subgroups and birth cohorts we examine.

We also 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 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 or lifespan, and correlations across persistence measures of SES and of health or lifespan are low.

Finally, we use our data to compute the first estimates of correlations in well-being that incorporate lifespan.² Our finding that mobility in lifetime utility is larger than mobility in SES alone echoes the results of Halliday et al. (2021), who compute correlations in measures of utility in adulthood based on income and health (rather than lifespan) and also document larger mobility in utility than SES. Thus, our findings suggest it is possible to improve mobility in well-being through channels other than economic mobility.

¹Two notable exceptions are recent papers by Kaplanis et al. (2018) and Minardi et al. (2024) who use non-representative sample of 130,000 and 350,00 individuals, respectively.

²Eshaghnia et al. (2022) compute lifetime well-being measures that use 14 different measures of income, consumption, and wealth. However, these measures do not incorporate lifespan.

II. Data and Descriptive Statistics

We exploit rich lifespan data from FamilySearch, a public wiki-style genealogical platform that includes over 13.7 million registered users and has profiles for over 1.7 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 attached to the person's profile (vital records, Censuses, etc.), and links to the profiles of their immediate relatives (parents, siblings, spouses, and children). Most profiles are created by individuals doing research on their own ancestors. FamilySearch provides an open edit format where anyone can make changes to any profile. The platform also works to enhance and verify information by attaching historical 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; Minardi et al. 2024).

To construct our estimation sample, we start with all individuals from the full-count US Censuses for 1900, 1910, and 1920. We focus on everyone in these Censuses who was born between the years 1880 and 1920, resulting in a base sample of 173.3 million person-year observations. FamilySearch provided us with a match file that indicates which of these individuals in the Census are attached to a profile on the FamilySearch website. We find that 133.3 million person-year observations from the Census are connected to a profile on FamilySearch (a 77% match rate). Since someone born before 1900 could have appeared in all three Census years, some individuals appear in multiple Census records. Our sample of 133.3 million person-year observations corresponds to 86.6 million unique profiles, of which 39 million include both a birth and death year for the individual.

Since most profiles on FamilySearch include connections to the individual's immediate family members, we use these connections to gather birth and death information for each individual's parents and siblings. Of the individuals in our sample for whom we have a death date, we match 77% to at least one parent with a death date, and match 67% to both parents. This results in a final sample that includes 26.1 million individuals for whom we have information on their lifespan and the lifespan of both of their parents. Unfortunately, we do not have cause of death information. Figure A.0 provides a flow chart that documents our sample size at each step of the data creation process.

The data creation process we use in this paper is distinct from the automated linking methods that are described in Bailey et al. (2020) and Abramitzky et al. (2021). Rather than using automated methods to link Census and death records to construct our sample, we are using the efforts of millions of people who have done family history research on FamilySearch. As such, the data we use in this paper is more similar to hand linked data created by Costa et al. (2017), genealogical data from Geni.com used by Kaplanis et al. (2018), or from Rootsweb.com used by Koylu et al. (2020). Data from genealogical websites is often described as the “gold standard” linked data and have been used as a benchmark for automated linkages (Bailey et al. 2020).

The parent-child relationships that we observe on the Family Tree are noted by individuals when creating the tree but often they are derived from Census records that include the individual living with their parents. However, parent-child relationships on the Family Tree can also come from birth, marriage, and death records, all of which often include information about an individual’s parents. Similarly, the information that we observe on the Family Tree about the individual’s birth date can come from Census records but often it is drawn directly from attached birth records or other records that include exact birth dates (rather than age as is reported in most Census years).

There are several reasons to believe that the linkages to birth and death information that we observe on the Family Tree are likely better than automated linking methods. First, we are able to link a Census record to a birth record by using information about the focal person and both their parents *and* by comparing the place of residence to the place of birth. Second, in our data, once the Census is linked to a birth record, the researcher can use the exact date of birth to better link to a death record. Third, in our data, we can link to a marriage record which can help in finding death info for women. Fourth, some death records include information about the family members (especially obituaries) which can help reduce false matches.

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.³ These national estimates are produced from state reports and try to account for missing populations: not all states

³ Available at <https://www.ssa.gov/oact/HistEst/CohLifeTables/2020/CohLifeTables2020.html>

are in the vital registration system (e.g. in 1900 only 10 states were in the death registration system, with 36 states covered in 1920), and there is under-registration in covered states (Hetzel, 1997). Although it is ultimately impossible to determine whether these data are representative of the nation, they constitute the gold standard.

To compare our sample with the SSA cohort life tables, we plot the raw lifespan histogram for our sample and for the SSA data beginning at age zero (see Figure A.1). We focus on the 1900 birth cohort. 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; Minardi et al. 2024).

However, if we condition on individuals surviving to 25 years old (the bottom panel of Figure A.1), the distribution of lifespan in our data is much closer to the distribution in the SSA data. Conditional on surviving to 25, the distribution of age at death is shifted to the right in our sample, with our sample's mean age at death slightly above that in the SSA data, but matching the shape of the distribution quite well.⁴ For the 1900 birth cohort, the expected age at death conditional on survival to age 25 for males in our data is 70.5, whereas it is 68.3 in the SSA data; for women, these figures are 76.7 and 74.6, respectively. The differences are smaller for the 1910 birth cohorts, for whom the gap is about a year, as shown in Table A.2. Part of this difference is likely because our final sample underrepresents Black individuals and immigrants, two groups with 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. The SSA sample is re-weighted to make it representative of the US, but the weighting may not be sufficient. Our data, on the other hand, includes all states.

b. Our Sample

There is a natural asymmetry in the lifespan distribution of our base sample and that of their parents stemming from the fact that not all individuals survive to become parents. As a result, we never observe what the correlation in lifespan would have been for individuals who did not live long enough to have children. To address this asymmetry and the issue of missing deaths among

⁴ This is also clear when comparing the survival curves for the 1900 and 1910 cohorts as shown in Figure A.2.

children, we restrict our main sample to parents and children who lived to at least 25, which is commonly done in studies of lifespan persistence (Table A.1) and is also what is typically done when investigating intergenerational persistence in socio-economic outcomes, which are not observable at younger ages. Thus, our study investigates adult lifespan, which we define as the age at death among those who survived to age 25. We test the robustness of our results to this choice of cutoff.

Our final sample includes 26.1 million children matched to both parents.⁵ This sample includes at least 30% of the target population (individuals in the 1900-1920 Censuses who were born between 1880 and 1920, survived to age 25, and whose parents also survived to age 25).⁶ 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. Second, because we use family trees to identify families, we are not limited to studying families living together at a given point in time, as would be the case if we tracked individuals from a given Census. Third, because the Family Tree aggregates 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 and common limitation in other datasets that obtain age at death from national databases such as the Death Mortality Files or the Numident. Finally, because our samples are very large, we can cut the data along multiple margins to study heterogeneity.

c. Summary Statistics

Table 1 presents summary statistics for our analysis sample. Our data include 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

⁵ 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.

⁶ There are 89 million individuals born from 1880 to 1920 observed in the 1900, 1910 and 1920 Censuses (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.

groups in the population: nearly all are white and very few are immigrants.⁷ Our data do have good representation of women, who are typically under-studied in analyses of intergenerational correlations (Hollingsworth 1976). Our data are also skewed towards the Midwest and under-represent 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 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, sex, 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 subgroups (e.g. Blacks or immigrants).

The average man in our sample (conditional on surviving to age 25) lived to age 70.2; the average woman lived 6 years longer, to age 76.1. This 6-year sex gap is consistent with previous research based on period data (Goldin and Lleras-Muney, 2019) and with the sex gaps in the SSA data, which are 6.2, 6.97 and 6.35 for the 1900, 1910 and 1920 birth cohorts surviving to 25. 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; 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.⁸ 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

⁷ 99% of individuals in our sample are white (instead of 86% in the 1880-1899 cohorts in the 1900 full census and 88% of the 1900-1920 cohorts in the 1920 full Census) and very few are immigrants (2% instead of 12% for the 1880-1899 cohorts and 1% instead of 2% in the 1900-1920 cohorts). (See Table A.3) It is well known that genealogies underrepresent Black individuals, immigrants, and children (Pope 1992).

⁸ Historical trends in life expectancy in the US are hotly debated because of data concerns – 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. 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. 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). 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. 2025).

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).

Starting in 1890, 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 in 1920 for women and hovered 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. 2025, Cutler et al. 2006).

The second notable trend is that the female advantage in lifespan grew substantially across these cohorts. This occurred in part because, for men, adult life expectancy was falling for cohorts born in the 1880s and only began rising modestly in the 1890s. But the gap also grew because, after 1900, the increase in longevity was larger for women than for men. 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.⁹

Figure 1 also shows the trends across cohorts using the SSA data, 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 do not suffer from large systematic biases and are broadly representative of the US population, despite the under-representation of smaller groups. Also notable is that our data do not appear to systematically overstate lifespan, a common issue in demographic studies (Hacker 2010).

III. A Simple Model of Lifespan and Implications for Mobility

We now present a simple model of lifespan to better understand intergenerational correlations in lifespan and their relation to correlations in SES and sibling correlations. Previous

⁹ 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). Other factors, such as lower exposure to infectious diseases likely contributed to these lifespan gains as argued by Goldin and Lleras-Muney (2019).

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 food, access to and quality of life saving technologies, the quality of social interactions, etc.), sex, SES (education, income and occupation), and random components (Cutler et al. 2006, van den Berg et al. 2017).

Let L_i be the lifespan (or age at death) of an individual i . Denote the sex of individual i as $s(i)$, where $s(i) = 1$ if i is male and zero otherwise. i lives in region $r(i)$ and is a member of birth cohort $c(i)$.

Suppose that the individual’s lifespan is determined uniquely by: their genes (G_i); environmental factors that are common to all individuals living in the same place and belonging to the same birth cohort ($\alpha_{r(i),c(i)}$); the individual’s socioeconomic status (SES_i); random factors (ε_i); and a within-family sex-specific random effect ($\gamma_i^{s(i)}$). This last term captures all biological, hormonal, behavioral and social factors that affect mortality differentially by sex within families. This term therefore goes beyond a simple male–female dummy: it can generate both the aggregate female–male difference in mean lifespan and the additional similarity of same-sex relatives documented in prior work (as we discuss below).

We start with a very basic model that treats these factors as linearly additive in the determination of lifespan. In this model, for individual i , lifespan is given by:

$$L_i = \delta G_i + \alpha_{r(i),c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i \quad (1)$$

where ε_i is a random error capturing the fact that lifespan may also be determined by unpredictable accidents. 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.¹⁰ The parameter λ represents the effect of SES on lifespan.

We make the strong assumption that all the determinants of lifespan (all the terms on the right hand side of Equation (1)) are uncorrelated with each other.¹¹ In addition to assuming linearity

¹⁰ 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 linear polygenic score to predict lifespan using the identified genes.

¹¹ This is clearly a strong assumption that is likely violated to some extent. For example, evolutionary biology argues 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) and empirical evidence has demonstrated that genes are indeed correlated with environments (e.g. Belsky et al. 2016).

in genes, sex-specific factors, SES and environment, Equation (1) also assumes that there are no interactions between sex-specific factors, genes and the environment, which have been shown to exist (e.g. van den Berg et al 2017).¹²

Equation (1) describes parents but also describes children. We now discuss how children’s and parents’ lifespans are related. We consider a nuclear two-generation family. Denote i ’s mother as $m(i)$ and denote i ’s father as $f(i)$. We assume for simplicity that siblings share the same mother and father and that a pair of parents share the same SES after forming a household.

First, we assume that children receive half of their genetic endowment from each parent so that $G_i = \frac{1}{2}[G_{m(i)} + G_{f(i)} + \eta_i]$, where η_i captures the fact that genes are transmitted with some noise (for example, due to mutations).¹³ We assume that η_i is the same for identical twins. We assume $G_{m(i)}$ and $G_{f(i)}$ are random variables with an unknown variance-covariance matrix reflecting the extent of assortative mating based on genetics among the parents. We assume that η_i is uncorrelated with all other terms. We are implicitly assuming that the variance of η_i is such that the variance of G_i is the same across generations.

It is worth noting here that the contribution of parental genes to a child’s lifespan is more complicated than it first appears. Although genes may determine lifespan, and parents pass on these genes to their children, each gene does not uniquely map onto specific traits. In fact, most traits, like height or lifespan, are polygenic – their expression depends on how hundreds of genes are combined. Thus, it is possible to have *traits* that depend highly on genetic endowments but that are only weakly transmitted from parents to children.

Second, we assume that parents and children may share common environments, either because they live in the same location or because they live in locations with similar characteristics. Thus, we allow α to be correlated across generations.

Next, we assume that the correlation structure in $\gamma_i^{s(i)}$ between mothers (fathers) and their children will differ by their sex. This is consistent with the fact that, empirically, lifespan correlations differ based on whether the parent and the child (or the siblings) share the same sex (see Table A.1). While it is unclear why lifespan correlations vary by sex within families, the

¹² This is also unrealistic. Life expectancy for women has grown more than that of men in the 20th century, which suggests environmental changes have favored females, see Goldin and Lleras-Muney (2019).

¹³ We assume that maternal and paternal genes contribute equally to child lifespan, which has yet to be shown in data.

previous literature provides several possibilities that are not clearly only genetic. These sex-specific correlations may occur for example because daughters (sons) will follow their mothers' (fathers') demographic behavior (e.g. age at marriage, number of children), their economic choices (e.g. their choice of occupation) and follow similar behavioral practices (e.g. smoking and drinking behaviors), all of which could affect lifespan and which varied over time and space but are typically highly gendered across societies.¹⁴ Specifically, we assume that $COV(\gamma_i^0, \gamma_j^1) = 0$ while $COV(\gamma_i^1, \gamma_j^1) \neq 0$ and $COV(\gamma_i^0, \gamma_j^0) \neq 0$ for any individuals i and j belonging to the same family.

We also allow SES to be correlated across generations, since, as noted earlier, a substantial literature has documented that SES is highly persistent within families.

Finally, we assume there are no intergenerational correlations across domains. For example, while we allow the SES of parents and children to be related, we assume that the SES of the father does not predict the location of his children (i.e. $COV(SES_{f(i)}, \alpha_{r(i),c(i)}) = 0$). While these correlations across domains may exist, we ignore them here for simplicity.

This data-generating model has several implications for absolute and relative mobility, sibling correlations, and for how the persistence in lifespan relates to the persistence in SES.

Absolute Mobility. For males, we define absolute mobility as $\Pr(L_i > L_{f(i)})$, namely the probability that a child lives a longer life than their same-sex parent (analogously, $\Pr(L_i > L_{m(i)})$ defines absolute mobility for females). If parents and children in a given family share the same SES (so SES is perfectly transmitted) and the same sex, live in the same location, and the environment is stationary (i.e. $\alpha_{r(i),c(i)} = \alpha \quad \forall r(i), c(i)$), then the expected lifespan of fathers (mothers) and sons (daughters) is identical, and so is their variance. For a closed population (where the joint distribution of G is constant), the only drivers of absolute intergenerational mobility will be changes in SES and the environment.

In general, however, the environment is not stable, as changes in public health policies and medical technologies impact lifespan over time. This suggests there will be differences in absolute

¹⁴ For example, for evidence on the sex-specific intergenerational transmission of fertility behavior, see Jennings, Sullivan and Hacker (2012); for evidence on the sex-specific intergenerational transmission of occupation, see Stevens (1986) and Hellerstein and Morrill (2011), and for evidence on the sex-specific intergenerational transmission of smoking and drinking, see Loureiro, van den Berg, and van Doorslaer (2010) and Handley and Chassin (2009).

mobility over time. In addition, the SES of a child could also differ from the SES of their parent due to broad economic shifts or through regression to the mean, in which case absolute mobility is determined by two factors: whether the environmental changes are affecting lifespan and whether children have a different SES than their parents. This will also be partly determined by migration, which will result in different environments and SES across generations.

Relative Mobility. Relative mobility measures capture a different concept: whether parents who live longer than their peers will have children who will live longer than their own peers. Our primary measure of relative persistence—which is the inverse of mobility—will be the regression coefficient of a child’s lifespan on the parents’ lifespan. This relative persistence measure (as well as alternative measures such as the Pearson correlation or the rank-rank slope) depends crucially on the covariance in lifespan between parents and children. For example, we can express the covariance between fathers and sons as:

$$COV(L_i, L_{f(i)}) = \delta^2 \frac{1}{2} COV[G_{m(i)}, G_{f(i)}] + \delta^2 \frac{1}{2} V(G_{f(i)}) + COV(\alpha_{r(i)c(i)}, \alpha_{r(f(i))c(f(i))}) + \lambda^2 COV(SES_i, SES_{f(i)}) + COV(\gamma_i^1, \gamma_{f(i)}^1) \quad (2)$$

This covariance, and thus our persistence measure, depends on genetic and non-genetic factors. There are three genetic components: how much genetic assortative mating there is on the part of the parents ($COV[G_{m(i)}, G_{f(i)}]$); how much parental genes affect a child’s lifespan (δ); and the variance of paternal genes. There are four other components: the extent to which parents and children share the same environment ($COV(\alpha_{r(i)c(i)}, \alpha_{r(f(i))c(f(i))})$); the extent to which parents and children share the same SES ($COV(SES_i, SES_{f(i)})$); the extent to which SES matters for lifespan (λ); and the covariance of the within-family sex-specific random effect ($COV(\gamma_i^1, \gamma_{f(i)}^1)$).

Sibling and Intergenerational Mobility. Our simple model has implications for how sibling coefficients (from regressing a child’s lifespan on their sibling’s lifespan) relate to intergenerational coefficients (from regressing a child’s lifespan on their parent’s lifespan), and how this depends on sex and on whether the siblings are twins, which we can test with our data. We can summarize the predictions of the model as stating that, if parents and children share the same SES and environment, the following will hold for males:

$$\beta_{twins} > \beta_{brothers} = \beta_{father-son} > \beta_{mother-son} > \beta_{father-mother} \quad (3)$$

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 (see Appendix A). Thus, the model predicts that (identical) twin correlations will exceed sibling coefficients, but sibling coefficients will be identical to parent-child coefficients within sex.

However, if the environment and SES siblings share is more similar than the environment and SES that parents and children share, then the sibling coefficients will exceed the parent child coefficients.¹⁵ It is unclear *a priori* whether father-son coefficients will exceed mother-daughter coefficients; this will depend on whether the covariance in the female-specific component is larger or smaller than the covariance in the male-specific component (i.e. whether $COV(\gamma_i^0, \gamma_j^0)$ or $COV(\gamma_i^1, \gamma_j^1)$ is larger within families). Similarly, it is unclear whether brother coefficients will exceed sister coefficients.¹⁶ Finally, although the simple model states that same sex coefficients will be larger than cross sex coefficients, this ignores possible differences by parental sex in the causal effects of parental lifespan on child lifespan: maternal lifespan could matter more than paternal lifespan for children of both sexes.

Relationship Between Intergenerational Persistence in Lifespan and SES. Prior research has estimated substantial persistence in SES across generations and has shown that SES is a strong predictor of lifespan. What does this imply for the intergenerational persistence of lifespan? Suppose that SES is generated as follows: $SES_i = \beta_{SES}SES_{f(i)} + \epsilon_i$, where β_{SES} is the intergenerational persistence in SES. We can then express the coefficient in lifespan $Coeff(L_i, L_{f(i)})$ (from running a regression of L_i on $L_{f(i)}$) as a function of the persistence in SES by modifying the expression in (2) to obtain (see appendix A):

$$Coeff(L_i, L_{f(i)}) = \lambda^2 \frac{var(SES_{f(i)})}{var(L_{f(i)})} \beta_{SES} + \pi \quad (4)$$

where π includes all other terms. This expression makes it clear that the contribution of β_{SES} is

¹⁵ In models with differential parental investment in children, sibling coefficients may also exceed intergenerational ones if parents aim to equalize outcomes among their children. (See Griliches 1979 and Behrman et al., 1982.)

¹⁶ The prediction that $\beta_{brothers} = \beta_{father-son}$ contrasts with 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 often 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 often a family background component that is not shared between parents and children.

likely small: λ (which is the effect of SES on lifespan) hovers around 0.3 for these cohorts (using education as a measure of SES, see Lleras-Muney et al. 2022); the variance of education is about 9.7 (Table 1); and the variance of lifespan is about 256. Altogether, then, $\lambda^2 \frac{\text{Var}(\text{SES}_{f(i)})}{\text{Var}(L_{f(i)})}$ is small ($\sim .0034$).

Lifespan and Lifetime Utility. To model the intergenerational transmission of well-being, we start with a simple model of lifetime utility. After normalizing the value of death to 0, the realized lifetime utility of a 25-year-old adult (U) is

$$U = \int_{t=25}^L e^{-\rho t} u(c_t) dt \quad (5)$$

where L is their lifespan (age at death), c_t is their consumption in year/age t , $u(\cdot)$ is the per period utility and ρ is their discount rate. Following Becker et al. (2005), we assume individuals receive income y_t every period and choose consumption to maximize expected utility subject to a lifetime budget constraint. If we assume, like they do, that per period income y is constant and that $\rho = r$, then it follows that the optimal consumption path consists of consuming one's income every period, so that $c_t = c = y$. Then we can rewrite equation (5) as

$$U = u(y) \int_{t=25}^L e^{-\rho t} dt$$

If we further assume that $\rho = 0$ so that people do not discount their future years, then realized lifetime utility is simply given by

$$U = u(y) * L_{25} \quad (6)$$

which states that the realized lifetime utility is simply the per period utility of income multiplied by the number of years the individual lives after age 25, L_{25} . Although this expression is based on a strong set of assumptions, it has the advantage of providing an intuitive measure that can easily be computed for all individuals with income and lifespan information. This model makes it clear why we care about the intergenerational transmission of lifespan: the longer people live, the higher their lifetime utility, independent of health.

IV. Absolute Lifespan 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 child's adult lifespan is on the y-axis and the parent's adult lifespan is on the x-axis. Lighter colors correspond to greater density. These joint distributions reveal several important patterns. First, the joint distribution is asymmetric: for all adults, there is a long tail of individuals who die relatively young, with substantially more compression at the top of the distribution. Second, while there is a positive correlation between parents' and children's adult lifespan, there is substantial variance. For example, consider men with fathers who 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% of men will die at either substantially younger or substantially older ages than their fathers.

These figures already suggest that a parent's adult lifespan is only weakly predictive of their children's adult lifespans. Indeed, some simple computations confirm what we see in the heat map. 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%.¹⁷ These shares do not vary meaningfully across cohorts. Most people live substantially longer or shorter lives than their parents.

Next, we investigate absolute mobility by assessing whether children live longer than their parents; this 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 absolute mobility will remain constant. There is in fact no mechanical relationship between increases in life expectancy and our main measure of absolute mobility.¹⁸

Figure 2a documents that absolute mobility in lifespan is substantially higher for women than men: most of the female joint distribution is above the 45-degree line, but most of the male joint distribution lies below the 45-degree line. Indeed, roughly 57% of women live longer than

¹⁷ 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.

¹⁸ 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 goes to war and dies 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.

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 measures for men are similar to existing estimates of absolute mobility in income for the 1880-1920 birth cohorts which were in the 40-50% range (Song et al. 2020; measures for women are not available).¹⁹

In this analysis, we require both children and parents to survive to age 25. However, this censoring rule is asymmetric in the sense that, if a parent is older than 25 when they have a child, then they have, by definition, survived to that (older) age, whereas we are requiring children to survive only to age 25. To test the sensitivity of our conclusions to this censoring rule, we implement an alternative rule that requires children to survive to the same age their parent was at time of their birth. This alternative censoring rule increases the absolute mobility estimates somewhat for both men and women, but the trends remain similar (see Figure A.5).

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. The mean difference in the ages at death of children and their parents (child-parent) is 3.8 for women and -1.4 for men (see Figure A.6). 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 again 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, medical technologies, or policies like health insurance expansions. 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 appear more mobile relative to high SES individuals because 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

¹⁹ 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).

(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, we see no other meaningful (and consistent) differences across groups defined by 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 between the longest-lived and the shortest-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, tracking cohorts over time. For the 1880-1900 birth cohorts, absolute mobility was increasing for women but decreasing for men. These trends reverse beginning around the 1900 birth cohort, with absolute mobility falling for women and increasing 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 A.6). 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 1880 and 1920 birth cohorts (Figures A.7) provides an explanation. There is a small group of daughters born in 1920 living substantially longer lives than earlier cohorts: 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 1900 or so. These increases explain why the average lifespan is rising. But there is also a relatively large group born 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.²⁰

²⁰ 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.7 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

They also highlight the importance of observing these joint distributions – and not just average lifespan – to understand trends in absolute mobility: it is only by observing how the distributions have changed in shape overtime that we can understand why mobility is decreasing.

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

V. Relative Mobility and Relative Persistence in Lifespan

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 is to compute transition matrices, which show the probability that a child born to a parent in a given quintile of the parental lifespan distribution ends up in a given quintile of the child lifespan distribution.

Table 2 shows that the transition matrices for both women and men are close to what we would expect if there were perfect mobility in lifespan: the diagonal elements are only somewhat larger than 20%, and the off-diagonal elements are not substantially smaller than 20%. Thus, there is substantial upward and downward mobility. For example, the son (daughter) of a father (mother) who was in the bottom quintile of lifespan has a 17.4% (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) conditional on having a parent who lived a long life (also in the top quintile) is about 25% for both fathers/sons and mothers/daughters, suggesting that long-lived parents have children who are only somewhat more likely to also be long-lived; this finding is consistent with those of van den Berg et al. (2019). This

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.

is the largest deviation from 20%, but even this deviation still suggests relatively low persistence, at least compared to what has been documented for income. 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%. Overall, relative mobility measures show high upward mobility. A convenient way to summarize the content of these matrices in a single statistic is to compute the “spectral gap” which corresponds to the difference in the highest eigenvalue (which is one) and the magnitude of the second highest eigenvalue. We obtain a value of 0.91 for both men and women, which is very close to 1, the value associated with perfect mobility (0 would indicate no mobility).

When we look across cohorts, we see that, in contrast to absolute mobility, relative mobility appears to be very stable across cohorts (Figure 5). Upward mobility (the fraction of children born to parents in the bottom 20% of adult lifespan who make it to the top 20th percentile of the distribution of adult lifespan in their cohort) did not change much, remaining at roughly 18% across cohorts. 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.

Together, these results show that absolute and relative measures of mobility give very different pictures of lifespan mobility. This is also true when looking at income (e.g. Chetty et al. 2017). In short, we again highlight the importance of both absolute and relative measures of mobility as noted by Deutscher and Mazumder (2023).

b. Relative Persistence

We now 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 is obvious from the heat maps, these slopes appear small. This is not because life expectancy is rising – lifespan increases across cohorts have no immediate or mechanical implications for intergenerational correlations or other measures of relative persistence such as regression coefficients. For example, if all children live 20 years longer than their parents (shifting the distribution up), then the correlation in lifespan across generations will remain the

same. Conversely, life expectancy could be stable, but the intergenerational persistence in lifespan could change if the joint distribution changes shape.

Our main specification relates the lifespan of the child (L_i) to the lifespan of either their father ($L_{f(i)}$) or mother ($L_{m(i)}$) by estimating equations of the type:

$$L_i = \beta_0 + \beta_1 L_{f(i)} + X\beta_2 + \epsilon_i \quad (7)$$

where L_i , our main variable of interest, is lifespan in years for individual i and $L_{f(i)}$ is the lifespan for individual i 's father, conditional on both the parent and child surviving to age 25. We also present results in logs and percentiles. In some specifications, we use the average of the parents' lifespan instead of the lifespan of one parent. 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 lifespan. 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 and assortative mating between the parents, as well as socio-economic influences and common environmental factors that affect both parents and children's adult lifespan as discussed earlier. 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 our simple model where there is a sex-specific component to lifespan. When we regress the lifespan of a son on the average of his parents' lifespan, we find a substantially larger coefficient of 0.141. This is consistent both with the fact that there is likely measurement error in our variable of interest (lifespan) as well as the fact that there is independent

information contained in each parent's lifespan (see Appendix A).²¹ Similarly, turning to daughters, we see that the daughter/father coefficient is 0.075 and the daughter/mother coefficient 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. We obtain nearly identical coefficients when we include parent and child birth-year fixed effects (Column 2), add state-of birth fixed effects for both parents and children (Column 3), and add controls for race and birth order (Column 4). While these characteristics have significant effects on the lifespan of the children in our sample, 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 lifespan percentiles are calculated within birth cohort (Column 2), and a log-log specification (Column 3). Our coefficient estimates are quite low across different functional forms. Raw correlations in lifespan are also smaller than our estimated regression coefficients (see Table A.4). For example, the raw father-son correlation is 0.08 (instead of 0.09) and the son-parent correlation is 0.10 (instead of 0.14). Thus, regardless of which specification we use, we find that the relative persistence of lifespan is always low. The coefficients are almost identical if we drop all deaths that occurred during WWI, WWII and the flu pandemic, confirming that low persistence is not the result of these uniquely deadly events (see Table A.5).

To check that our linear specification is appropriate, we plot the relationship between the lifespans of parents and children (see Figure A.8). We do this both for the lifespan and the lifespan percentile. We find that, for lifespan, the relationship is flat for those whose parents died before age 40 and then becomes steeper, though it 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 percentiles, the relationship is linear at almost all percentiles except for the very bottom and the very top.

Up to this point, we have conditioned on the child surviving to age 25. Prior research has used different age 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

²¹ The correlation between father and mother adult lifespan is quite low, at only 0.05 (see Table A.4).

effects, it is more appropriate to condition on living to very old ages, as younger deaths are more likely due to accidents. However, we find that coefficients are remarkably robust to the choice of cutoff ages (which we allow to vary from 0 to 65), remaining low across all specifications (see Figure 6).²² These results suggest that the genetic contributions to lifespan are small, an issue we return to later. We also estimate almost identical coefficients if we choose a different censoring rule, where we condition on children living to the age of the parent at the time of the child's birth. For example, the father son coefficient using age 25 as the cutoff is 0.09 and it is 0.087 if we use the symmetric cutoff rule instead (see Table A.6).

While the overall lifespan persistence 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 persistence of lifespan. However, persistence in lifespan is similar by parent's characteristics, child characteristics, and place of birth (Figure 3b). Even when we do observe differences, the range of lifespan persistence is always between 0.06 and 0.10, suggesting low persistence for all groups. This is true whether we use lifespan levels or ranks. 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 lower 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. Trends in Relative Persistence

Figure 7 shows the intergenerational persistence in lifespan in levels by year of birth of the child and by sex (Panel a). For men, we find that the lifespan persistence increases from around 0.06 to around 0.10 over the first two decades of the sample before plateauing for the 1900 to 1920 birth cohorts. Women see a similar increase for those born in the first two decades but then see a decline for those born in the last two decades to roughly 0.07. The lifespan persistence between

²² In this figure we also report the estimates that do not restrict the child's age to age 25. We do not find this matters, but as we note in the text individuals under age 25 are very poorly represented in our data, so these results are only suggestive.

children and the lifespan of both parents increases from around 0.10 to 0.15 for the first 20 birth cohorts before also plateauing. While these increases are large in a relative sense, the magnitudes remain low.

While somewhat muted, a similar pattern appears when using a rank-rank specification (Panel b). While the lifespan persistence 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 lifespan persistence 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 very different stories. Interestingly, research has shown that the rank-rank correlation in occupational status also remained relatively constant (around 0.3 to 0.4) for men born 1880-1920 (Song et al. 2020, Ward 2023, Buckles et al. 2023) and women born in this period (Buckles et al. 2023).

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 cohort 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)

VI. Mobility in Lifespan, Socio-Economic Status and Lifetime Well-being

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 cohorts and data selection criteria. As a result, it is unclear whether these differences are due to true differences in persistence across domains or differences in the samples used in each study.

To make progress on this issue, we compare correlations in SES and lifespan in two samples. First, we follow the economic history literature and use occupation-based incomes observed in the Census to construct a sample of fathers and sons for whom we observe both

occupation and lifespan. Specifically, we use the occupation score (“occsore”) variable available in the IPUMS data which assigns to each individual the median income in 1950 for their occupation. Second, we look at sibling correlations. The sibling correlations are computed using the 1940 Census which has occupation-based income and also, crucially, reports education and household incomes. As noted earlier, in our simple model of lifespan, if there are significant economic and environmental changes (which is the case for the cohorts we study), siblings likely grow up in a more similar environment than parents and children—as siblings share the same parents, home, and neighborhood—in which case sibling correlations will exceed the measure of intergenerational persistence. This has been shown to be true for measures of SES; Björklund and Jäntti (2012) use data from Sweden and estimate brother correlations in schooling of 0.46 and father-son correlations of 0.39; similarly, the brother correlations in earnings are 0.24 while father-son correlations are 0.14.

a. Intergenerational Persistence in Lifespan and SES for Men

We create a sample for this exercise by taking the data used by Buckles et al. (2023) who compute persistence in occupation-based incomes for the 1880, 1890, 1900, 1910 and 1920 birth cohorts. We then keep the subset of fathers and sons for whom we also observe lifespans. We focus on men (as the previous literature has done), as women’s labor force participation rates were low in this period.

Figure 8 presents the results. The yellow line plots the coefficients that result from regressing the occupation percentile of the son on the occupation percentile of the father. This coefficient is 0.26 for the 1880 cohort, rises to 0.35-0.36 for the 1890 and 1900 cohorts, and rises again somewhat to 0.39 for the 1910 cohort. Although our sample is smaller and potentially selected, these estimates closely mirror the levels and trends documented in Buckles et al. (2023). The bottom line shows the estimates of the coefficients for lifespan and shows that, in this sub-sample, we also closely replicate our earlier findings: coefficients rise from 0.06 for the 1880 cohort to about 0.1 for cohorts born after 1890. Overall, the SES coefficients are about 4 times larger than the coefficients for lifespan, confirming the intuition from comparisons to the literature: SES persists at a much higher rate than lifespan.

b. Sibling Correlations in Lifespan and SES

The previous results have two main limitations. First, they only include men, and second, they rely on occupation-based income as a measure of SES. To verify our conclusions, we construct a subsample of sibling pairs for whom we observe lifespan and whose education, occupation, and income we also observe in the 1940 Census. We construct this sample by keeping individuals in our data who have the same two parents and who are linked to the 1940 Census. We restrict the sample further to include only those with non-missing information on all the SES variables in addition to lifespan. This sibling sample is observably similar to the full analysis sample though somewhat longer lived (see Table A.7).

We regress each child's outcome on the outcome of their same-sex sibling or opposite-sex sibling and include birth cohort fixed effects for each sibling (Table 5). We include all unique possible sibling pairs as observations and cluster standard errors by family. We estimate sibling coefficients for adult lifespan, education, and earnings (both individual and household) using a constant sample with non-missing data on all outcomes (Columns 1-4). Because many women do not have positive earnings in the 1940 Census, we also consider a less restricted sample, which includes all individuals with non-missing measures of education and lifespan (but no restriction on earnings) as a robustness check (Columns 5 and 6).

Several conclusions emerge from these sibling associations. First, similar to Minardi et al. (2024) and in line with most prior studies (Table A.1), the coefficients for lifespan among siblings are larger than our parent-child intergenerational correlations: for example, the brother-brother coefficient is about 0.13 and the father-son coefficient is 0.08 for this same sample (shown in the last column).

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 similar coefficient for male siblings' adult lifespan (0.13) or female siblings' (0.11) but a rather weak relationship between brothers' and sisters' lifespan (0.04). This is consistent with our model's predictions.

Third, our estimates for SES similarity across siblings are consistent with those in the existing literature, albeit a bit larger, suggesting our sample may be broadly representative. The coefficient for education is 0.55 for brothers, 0.60 for sisters, and 0.53 for opposite-sex siblings. When we consider correlations instead of regression coefficients, we estimate correlations of 0.55 for brothers, 0.59 for sisters, and 0.53 for mixed-sex siblings (see Table A.8). These are broadly

consistent with the literature, which has 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 larger for women: Solon et al. (1991) estimate a 0.28 correlation for the 1951-1958 birth cohorts, and Mazumder (2008) estimates a 0.34 correlation for the 1947-1955 cohort. A similar picture emerges when examining income. We estimate a coefficient of 0.25 for males and 0.16 for females, comparable to those estimated in the literature, which generally range from 0.1 to 0.4 using a single year of earnings.²³

One puzzling finding is the negative sister/brother coefficient on income; 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. The coefficients estimated using household income are similar for sisters (0.36), brother-sister pairs (0.33), and brothers (0.35).

Fourth, we see much larger coefficients (and correlations) for education or income than for lifespan within this constant sample. 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 coefficients 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. These findings confirm our earlier observations that persistence in SES is substantially larger than persistence in lifespan, and this holds true regardless of the measure of SES we employ.

c. 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 across cohorts. Although lower in levels, one might expect that intergenerational correlations in adult lifespan are driven by the same factors that make SES persistent. If this is the case, then in places where education or income are more persistent from one generation to the next, lifespan would be more persistent as well. Persistence in education,

²³ See Solon (1999). Correlations are in fact higher when more permanent measures of income are computed (Eshaghnia et al. 2023).

income and health might also be high in the same locations or periods because education, income, and adult health are highly correlated.

We first estimate this relationship for the subsample of men for whom we observe both occupation-based income percentiles and lifespan. In the left panel of Figure 9 we plot the estimated lifespan coefficients (from regressing children's lifespan on parental lifespan) on the y-axis against the estimated coefficients for occupation-based income percentiles for each birth cohort on the x-axis. The estimated relationship (using the inverse variance of the estimated coefficients for occupation-based income percentiles as weights) is slightly negative and statistically insignificant.²⁴ In the right panel, we estimate these coefficients by state of birth instead. Again, the relationships are slightly negative and statistically insignificant. Surprisingly, we do not find that the persistence of SES and of lifespan are related.²⁵

To confirm these results for both sexes while using better measures of SES, we conduct a similar exercise using our sibling sample. 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. As a result, for this analysis, we restrict our sample to siblings who survive until at least 60 years old so that all cohorts are conditioned on living until the same age and can be compared. This is unlikely to affect our conclusions, given that we observed similar measures of persistence regardless of our choice of age cutoff.

In Figure 10, we plot the sibling coefficient on lifespan (y-axis) across states (or across cohorts) against the sibling coefficient on education (x-axis). Again, when we look across states (right panels), we see that sibling SES coefficients and sibling lifespan coefficients are not systematically related; places where the coefficient in education is large are *not* the same places where the coefficient in lifespan is high. The same is true across cohorts (left panels). A simple regression finds that the slopes summarizing these relationships are very small and, in most cases, statistically insignificant, and this is unchanged by using the estimated variances as weights.²⁶

²⁴ The results are very similar if we do not weight the estimates.

²⁵ These regressions may underestimate the relationship between persistence in lifespan and persistence in education or income because we are using aggregate correlations as data. We can estimate these instead using individual data by defining the dependent variable of interest as the gap in lifespan between a child and their parent, and the independent variable as the gap in SES between a child and their parent. The correlation in father-son gaps across domains we obtain is small, at about 0.02. The regression coefficient when we regress the gap in lifespan on the gap in SES is -0.009 and is statistically significant at the 5% level. In all cases these relationships are very small (see Table A.9).

²⁶ These results are very similar if we regress gaps in lifespan across siblings on gaps in SES after computing the gaps using the individual level data (see table A.10). We find that the brother/brother or sister/sister correlations in gaps is

Surprisingly, we do find a *negative* and statistically significant coefficient for brothers and sisters, indicating that when education correlations are high among different sex siblings, they tend to be low for lifespan. It is not clear why this is the case.

In sum, we find that the persistence of adult lifespan and of socioeconomic status are unrelated, whether we measure this using intergenerational persistence in occupation and lifespan or sibling coefficients for education and lifespan. This highlights that families and communities that succeed 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.²⁷

d. Computing the Intergenerational Persistence in Lifetime Well-Being

The fact that persistence of SES and the persistence of adult lifespan are independent of each other implies that mobility in well-being may be different from what economic measures alone would suggest. We now turn to computing persistence in well-being using lifetime utility as our well-being measure.

We start by computing lifetime utility following equation (6). Once we choose a functional form for $u(y)$ and calculate lifetime utility for an individual, we can regress the lifetime utility of the child on the lifetime utility of the parent. We consider two alternatives for $u(y)$. The first is to assume that $u(y) = \sqrt{y}$, as is common in the literature, which assumes there are decreasing returns to consumption/income. To compute this in a consistent way across generations, income needs to be appropriately deflated. A second approach is to assume that $u(y) = \text{percentile}(y)$, where the percentile denotes a person's income percentile in their cohort.²⁸ This functional form is more unusual, as it assumes that individuals only derive utility from their position in the income

very small and so are the regression coefficients. Interestingly the only correlations and coefficients that are not small are the ones for opposite sex siblings which are negative, similar to what we find using the aggregate data.

²⁷ Fletcher and Jaitner (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 too find that the two measures are only weakly related.

²⁸ Recall the utility is relative to the value of death, which has been normalized to 0. In principle, either utility function could also include a constant representing the value of being alive – a positive constant would make the variation in lifespan relatively more important than the variation in income. While it is possible to calibrate the value of this constant (see Hall and Jones 2007 or Becker et al. 2005), we would require additional data to know how individuals tradeoff income versus survival probabilities.

distribution, rather than from their income levels – individuals likely care about both. However, by using percentiles, we avoid adjusting for inflation or accounting for lifecycle differences. Moreover, we can compare our results to the literature, which has often used rank-rank correlations to estimate the intergenerational persistence in income.

To compute these measures, we need to observe income (or income percentile) and lifespan for both parents and children, so we use the sample of men we constructed in Section VI a, where we observe occupation-based income and lifespan for both generations. Figure 8 shows the results. The red line shows estimates for the intergenerational persistence in lifetime utility. These coefficients are larger than the lifespan coefficients, but smaller than the income coefficients, showing that, at least for men, incorporating lifespan reduces the persistence of well-being below that of SES. The figure also shows that there is a bigger increase in the persistence of lifetime well-being over time than there is in either income or lifespans.²⁹

In sum, lifespan associations are both small in levels and small relative to SES associations. They also appear to be nearly orthogonal from SES associations. Lifetime utility measures which combine SES and lifespan outcomes into a single measure of well-being show less persistence than SES and exhibit different trends than persistence in each of its components.

VII-Why is Lifespan Persistence So Low?

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 intergenerational persistence in lifespan 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

²⁹ To test if our estimates differ using actual household income measures, we turn again to the sample of siblings, which we further restrict to those that also have occupation reported in 1940 (see Table A.11). Columns 1 and 3 demonstrate that our original results hold in this smaller sample, suggesting it is representative. We also compute sibling associations in lifespan utility. We find that this coefficient is ~ 0.42 if we use occupation percentiles to compute lifetime utility (column 4) or ~ 0.47 if we use household income instead (column 5). These coefficients are much larger than the coefficient for each individual component, which might be unintuitive. However, the coefficient for lifetime utility is not constrained to be between the coefficients of its components. It also depends on for example on how the parent's SES affects the child's lifespan (see Appendix derivations). Another reason we might see this pattern is if parents aim to equalize the lifetime utility of all their children (as argued by Griliches 1979 and Behrman et al., 1982). According to this type of model, parents might invest in the education of some and the health or marriage prospects of others. Prior research suggests that indeed parents treat children differently across domains (e.g. Grätz, M., & Torche, 2016; Breinholt and Conley 2023). If this is true, then summary measures of well-being that combine outcomes across different domains, like household income and lifetime utility, will be more correlated among siblings than across generations, which is precisely what we observe.

11 plots the estimates from a few of these studies and compares them directly to ours (where we estimate an 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 cohorts, 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 which are reflected in our simple model. 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.³⁰ Third, behaviors that determine health and SES are also passed 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, meaningfully overstate the level of upwards and downward mobility. We investigate these issues now.

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 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 (see Table A.13). Despite the smaller sample size, the intergenerational persistence in lifespan remains low and very similar in absolute value. To rule out that this is the result of low quality in the father's death certificate information (which might be a concern given that, in the US, vital records were not well kept until the 1930s), we also compare sibling correlations. We see a small (e.g. roughly 11-17% in Panel A) increase in the

³⁰ In Table A.12, we show that, in our data, education, income and occupation are significant predictors of lifespan.

correlations in the sample with certificates, suggesting measurement error is relatively small and that correcting for it would not meaningfully change the main conclusions of our paper.³¹ Thus, the persistence in lifespan may be somewhat underestimated due to measurement error, but this error is not sufficiently large to produce measure of persistence in the range of those for SES we observe.

b. Bounding the Contribution of Genes

To better understand the role of genes in the intergenerational transmission of lifespan, we leverage 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 persistence in lifespan to be higher among identical twins than among siblings. In contrast, if genes are less important (or unimportant), we would expect the persistence in lifespan to be similar among siblings and identical twins.

We have a large sample of twins – more than 100,000 pairs of twins with information on adult lifespan.³² Table 6 shows our 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 coefficients (0.10). This is true even though the association with father’s lifespan is almost identical among siblings and twins.

Unfortunately, we cannot differentiate between identical and fraternal twins, which means our estimates still understate the role of genetics; however, since we know the share of same-sex twins who are identical in the population, we can adjust our estimates accordingly. Since approximately 50% 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 coefficients, to determine the coefficients for identical twins.³³ 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 75% to 120% 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,

³¹ We also tried restricting attention to samples in which both parents and children have exact birth and death information (a measure of information quality) and for whom there are at least two sources attached to the tree documenting birth and death events with very little change in the results.

³² We identify twins as siblings who are born to the same parents in the same year and month.

³³ For this computation we assume that the variance in lifespan is identical in the twin and sibling samples.

since identical twins share the same genetics, while parents and children share only half. But twins also have more similar home environments relative to non-twin siblings. The identical twin coefficients in lifespan are still below 0.25, suggesting that the genetic heritability of lifespan is relatively low (at least relative to other traits like height).

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).³⁴ 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.³⁵ Although this may seem surprising, a possible reason why genetic influences on lifespan are small is that there is little evolutionary pressure to transmit genes that increase lifespan beyond reproductive years (Lee 2003).

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

Another 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, in which 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.

³⁴Closely related is work by Hjelmberg 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.

³⁵The 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 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).

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, following Björklund and Salvanes (2011).

Table 7 presents the results for our sibling sample. Among all siblings, the correlation in lifespan (which also corresponds to the R-squared with just family fixed effects in a one-way fixed effects model) 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 adjusted R-squared of the regression is quite small (0.039); when we estimate the same specification for education, the adjusted R-squared of the regression is 0.13 (Panel B).³⁶ When we include family fixed effects, the adjusted R-squared increases substantially for education (to 0.59) but much less so for lifespan (0.131), demonstrating that families have a much 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 intergenerational persistence in lifespan is low because the stochastic component of lifespan is large, and larger than for health or SES outcomes.

The intuition for this result is easily seen if we use a simplified version of how lifespan is determined. Suppose that the lifespan of an individual i is given by $L_i = H_i + e_i$, so that there are only two elements that matter: a person's health and a stochastic shock, which we assume to be orthogonal to H . Suppose further that the parent and the son have exactly the same H (H is perfectly transmitted across generations) so that the correlation in H is exactly equal to one, but their lifespans are not equal because each of them receives a different shock e (drawn from the same distribution). If we regress the lifespan of the child on the lifespan of the parent then the coefficient from this regression will be given by $V(H)/[V(H) + V(e)]$ and it will not be equal to 1, it will suffer from "attenuation" (compared to the coefficient for health), and this attenuation is larger the larger the variance of the error term in lifespan $V(e)$.

The coefficient we estimate is not biased in a traditional sense, however, because we are describing the persistence in lifespan across generations, not trying to uncover the persistence in health. As our model of welfare suggests, lifespan itself is an important input into welfare because

³⁶ The controls include birth cohort of mother FE, birth cohort of father FE, child cohort FE, place of birth FE, indicators for race, sex, number of siblings, birth order, mother and father immigrant status.

it determines the number of periods that can be enjoyed independently from the health enjoyed in a given period. But the example illustrates that even if health or other determinants of lifespan that the family transmits like SES are highly correlated across generations, lifespan can be much less correlated across generations if the variance of the stochastic term that determines lifespan is large. Our fixed effects results demonstrate that this is indeed the case.

VIII. Conclusion

In this paper, we use newly available data from family trees that include over 26 million individuals born in the United States from 1880 to 1920, and their parents, to estimate the intergenerational transmission of lifespan. We find that absolute mobility in lifespan – the probability that a child lives longer than their same-sex parent – was large, but larger among women than men, and it trended in opposite directions for men and women in the cohorts we study. Relative mobility was also high. 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. Persistence is low across subgroups and across cohorts, which provides insight into why earlier, very different studies – often conducted with relatively small convenience samples – come to similar conclusions.

The consistency of estimates across studies, samples, and cohorts implies that the results on relative persistence we obtain for the 1880-1920 cohorts may be applicable to cohorts born more recently. However, while persistence is low for all, there is substantial variation in these estimates across subgroups and across cohorts in *relative* terms. For example, the persistence of lifespan is 40-70% larger for those with high education. While these results suggest there may be notable heterogeneity, future work will need to carefully examine the extent to which the subgroup samples are representative and to examine why these estimates differ by SES. Nevertheless, since our measures of persistence are low for all groups, and our measures of absolute mobility are large, our results suggest that information about parental ages at death is not particularly informative about the age at death of an individual. Future research could investigate whether causes of death (which we do not observe) are more correlated across generations than ages at death.

We also find that lifespan correlations are low (and much lower than SES correlations). We show this is partly 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.

Finally, we document that incorporating lifespan into measures of welfare has large impacts on the intergenerational transmission of well-being: lifetime utility is less persistent than income across generations. An important direction for future research will be to better understand how different intergenerational measures are related, and to further 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 lifespan, 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 born in the 1880s lifespan declined, whereas lifespan rose very substantially for women born 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 and the need to better understand the sources of their divergent paths.

II. References

- Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., and Perez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, 59(3), 865-918.
- Abramitzky, R., Boustan, L., Jacome, E., and Perez, S. (2021). Intergenerational Mobility of Immigrants in the United States over Two Centuries. *American Economic Review*, 111 (2): 580-608.
- 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.
- Andersen, C. (2021). Intergenerational health mobility: Evidence from Danish registers. *Health Economics*, 30(12), 3186-3202.
- Bailey, M. J., Cole, C., Henderson, M., and Massey, C. (2020). How well do automated linking methods perform? Lessons from US historical data. *Journal of Economic Literature*, 58(4), 997-1044.
- 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.
- 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.
- Behrman, J. R., Pollak, R. A., & Taubman, P. (1982). Parental preferences and provision for progeny. *Journal of political economy*, 90(1), 52-73.
- 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, H., S, E. Black, A. Lleras-Muney and J. Price (2025). The Evolution of Lifespan Among Cohorts Born at the Turn of the 20th Century US. Mimeo.
- Björklund, A., & Jäntti, M. (2012). How important is family background for labor-economic outcomes?, *Labour Economics*, 19(4), 465-474.

- Björklund, A., & Jantti, 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, S. E. & Devereux, P. 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.
- Breinholt, A., & Conley, D. (2023). Child-driven parenting: differential early childhood investment by offspring genotype. *Social Forces*, 102(1), 310-329.
- 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.
- 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., 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.
- Costa, D.L., DeSomer, H., Hanss, E., Roudiez, C., and Wilson, S. (2017). Union Army Veterans, All Grown Up. *Historical Methods*, 50(2), 79-95.
- 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.
- 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.
- Eshaghnia, S., Heckman, J. J., Landersø, R., & Qureshi, R. (2022). *Intergenerational transmission of family influence* (No. w30412). National Bureau of Economic Research.
- Eshaghnia, S., Heckman, J. J., & Landersø, R. (2023). *Maximum Impact Intergenerational Associations* (No. w31093). National Bureau of Economic Research.
- 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. (2023). Understanding Geographic Disparities in Mortality. *Demography*;60(2):351-377. doi:10.1215/00703370-10609710.
- Fogel, R. W. (1986). Nutrition and the decline in mortality since 1700: Some preliminary findings. In *Long-term factors in American economic growth* (pp. 439-556). University of Chicago Press.
- 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.
- Grätz, M., & Torche, F. (2016). Compensation or reinforcement? The stratification of parental responses to children's early ability. *Demography*, 53(6), 1883-1904.

Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *Journal of political Economy*, 87(5, Part 2), S37-S64.

Hacker, J. D. (2010). Decennial life tables for the white population of the United States, 1790–1900. *Historical methods*, 43(2), 45-79.

Hall, R. E., & Jones, C. I. (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics*, 122(1), 39-72.

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.

Handley, E. D., & Chassin, L. (2009). Intergenerational transmission of alcohol expectancies in a high-risk sample. *Journal of Studies on Alcohol and Drugs*, 70(5), 675–684.

Hellerstein, J.K., & Morrill, M.S. (2011). Dads and Daughters: The Changing Impact of Fathers on Women's Occupational Choices. *Journal of Human Resources* 46(2), 333-372.

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

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

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.

- Jennings, J. A., Sullivan, A. R., & Hacker, J. D. (2012). Intergenerational transmission of reproductive behavior during the demographic transition. *The Journal of interdisciplinary history*, 42(4), 543–569. https://doi.org/10.1162/jinh_a_00304
- 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.
- 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).
- Kletzky, O. A., Nakamura, R. M., Thorneycroft, I. H., & Mishell, D. R. (1975). Log normal distribution of gonadotropins and ovarian steroid values in the normal menstrual cycle. *American Journal of Obstetrics and Gynecology*, 121(5), 688-694.
- Koylu, C., Guo, D., Huang, Y., Kasakoff, A., & Grieve, J. (2020). Connecting family trees to construct a population-scale and longitudinal geo-social network from the US. *International Journal of Geographical Information Science*, 35(12), 2380-2423.
- Kumar, S., & Nahlen, B. (2023). Intergenerational persistence of health: Evidence from India. *Economics Letters*, 224, 111023.
- Lee, R. D. (2003). Rethinking the evolutionary theory of aging: transfers, not births, shape senescence in social species. *Proceedings of the National Academy of Sciences*, 100(16), 9637-9642.
- 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.

- Loureiro, M. L., van den Berg, G. J., & van Doorslaer, E. (2010). Smoking habits: Like father, like son; like mother, like daughter. *Oxford Bulletin of Economics and Statistics*, 72(2), 297–316.
- Lu, F., & Vogl, T. (2023). Intergenerational persistence in child mortality. *American Economic Review: Insights*, 5(1), 93-109.
- Manski, C. F. (2011). Genes, eyeglasses, and social policy. *Journal of Economic Perspectives*, 25(4), 83-94.
- 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.
- Minardi, S., Corti, G., & Barban, N. (2024). Historical patterns in the intergenerational transmission of lifespan and longevity: A research note on US Cohorts Born between 1700 and 1900. *Demography*, 61(4), 979-994.
- 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 Genetics*, 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.
- 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.
- 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., & Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography*, 43(4), 631-646.
- Solon, G. (1999). Intergenerational mobility in the labor market. In *Handbook of labor economics* (Vol. 3, pp. 1761-1800). Elsevier.
- 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.
- Stevens, G. (1986). Sex-Differentiated Patterns of Intergenerational Occupational Mobility. *Journal of Marriage and Family*, 48(1), 153–163.
- Stuhler, J., & Biagi, F. (Eds.). (2018). *A review of intergenerational mobility and its drivers*. Luxembourg: Publications Office of the European Union.
- 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.
- Travison, T. G., Vesper, H. W., Orwoll, E., Wu, F., Kaufman, J. M., Wang, Y., ... & Bhasin, S. (2017). Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the United States and Europe. *The Journal of Clinical Endocrinology & Metabolism*, 102(4), 1161-1173.
- 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-Gironde, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A., ... & Slagboom, P. E. (2019). Lifespan defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature communications*, 10(1), 35.
- Vaupel, J. W. (1988). Inherited frailty and lifespan. *Demography*, 25(2), 277-287.
- 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 (16.08)	70.2 (15.43)	76.1 (16.25)
Father's Average Lifespan	71.7 (13.55)	71.7 (13.53)	71.7 (13.58)
Mother's Average Lifespan	72.3 (15.89)	72.3 (15.84)	72.3 (15.96)
Father's birth year	1867 (14.09)	1867 (14.05)	1867 (14.15)
Mother's birth year	1872 (13.30)	1872 (13.25)	1872 (13.35)
Birth year	1901 (11.62)	1901 (11.62)	1901 (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 (2.36)	2.87 (2.35)	2.91 (2.37)
Birth Order	2.39 (1.68)	2.39 (1.68)	2.40 (1.69)
Mother's Age at Child's Birth	29.1 (6.71)	29.1 (6.69)	29.2 (6.73)
Father's Age at Child's Birth	33.9 (8.02)	33.9 (7.99)	34.0 (8.05)
Education	9.57 (3.12)	9.45 (3.21)	9.70 (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, were successfully matched to the Family Tree, and for whom we could compute age at death. See text for further details on sample construction and sample selection. As education is only available in the 1940 census onward, we

know the years of education achieved for only a subset of our sample. There are 15,335,985, 8,260,372, and 7,075,613 observations with education for the full sample, sons, and daughters respectively.

Table 2. Lifespan Quintile Transition Matrix, by Sex

		Mother Quintile				
		Bottom				Top
		1	2	3	4	5
Daughter Quintile	1	22.40	22.24	20.34	18.59	16.43
	2	20.90	21.63	20.71	19.24	17.07
	3	20.22	20.57	20.91	20.62	19.31
	4	18.99	18.84	19.90	20.90	21.58
	5	17.50	16.72	18.15	20.66	25.61

		Father Quintile				
		Bottom				Top
		1	2	3	4	5
Son Quintile	1	22.87	22.27	20.46	18.75	16.57
	2	21.29	21.51	20.62	19.40	17.41
	3	19.77	20.06	20.30	20.06	19.19
	4	18.68	18.84	19.81	20.71	21.41
	5	17.38	17.33	18.81	21.08	25.43

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. These matrices compare the portions of the son/father (mother/daughter) sample in a lifespan quintile given their father's/son's (mother's/daughter's) lifespan quintile. N = 13,944,386 for men and 12,189,775 for women. Q1 refers to the 20 percent of individuals that have the lowest lifespan, and Q5 refers to those with the highest.

Table 3. Intergenerational Persistence in Lifespan for Varying Child and Parent Pairings and Specifications

Model	Outcome: Lifespan (Years)				# of Obs.
	(1) Lifespan (Years) No Controls	(2) (1) + Parent and Child Birth Year FE	(3) (2) + Parent and Child State of Birth FE	(4) (3) + Race and Birth Order Dummies	
Son/Father	0.089 (0.0003)	0.090 (0.0003)	0.087 (0.0003)	0.087 (0.0003)	13,944,386
Son/Mother	0.062 (0.0003)	0.062 (0.0003)	0.059 (0.0003)	0.059 (0.0003)	13,944,386
Son/Parents' Average	0.141 (0.0004)	0.141 (0.0004)	0.137 (0.0004)	0.137 (0.0004)	13,944,386
Daughter/Father	0.075 (0.0004)	0.075 (0.0004)	0.072 (0.0004)	0.072 (0.0004)	12,189,775
Daughter/Mother	0.081 (0.0003)	0.073 (0.0003)	0.071 (0.0003)	0.071 (0.0003)	12,189,775
Daughter/Parents' Average	0.150 (0.0005)	0.142 (0.0005)	0.138 (0.0005)	0.137 (0.0005)	12,189,775

Notes: Each cell separately provides the estimated regression coefficient in lifespan between the two individuals indicated in the row label. Errors are clustered at the family level. 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. Intergenerational Persistence in Lifespan for Varying Child and Parent Pairings and Measures

Model	Outcome			
	Lifespan (Years)	Percentile	Log Lifespan	# of obs.
Son/Father	0.090 (0.0003)	0.090 (0.0003)	0.076 (0.0003)	13,944,386
Son/Mother	0.062 (0.0003)	0.078 (0.0003)	0.048 (0.0003)	13,944,386
Son/Parents' Average	0.141 (0.0004)	0.162 (0.0004)	0.132 (0.0004)	13,944,386
Daughter/Father	0.075 (0.0003)	0.079 (0.0003)	0.059 (0.0004)	12,189,775
Daughter/Mother	0.073 (0.0003)	0.094 (0.0003)	0.056 (0.0003)	12,189,775
Daughter/Parents' Average	0.142 (0.0004)	0.166 (0.0004)	0.127 (0.0005)	12,189,775

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

Table 5. Sibling Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample:	Individuals with data on lifespan, education, individual income and household income				Individuals with data on lifespan and education		Individuals with data on lifespan
Outcome:	Adult lifespan	Education	Income	HH Income	Adult lifespan	Education	Father-child lifespan
Brother/Brother	0.134 (0.0006) 3,663,633	0.554 (0.0007) 3,663,633	0.251 (0.0138) 3,663,633	0.345 (0.0043) 3,663,633	0.134 (0.0006) 4,125,511	0.552 (0.0007) 4,125,511	0.084 (0.0005) 4,679,398
Sister/Sister	0.106 (0.0007) 2,402,118	0.602 (0.0008) 2,402,118	0.157 (0.0101) 2,402,118	0.357 (0.0048) 2,402,118	0.106 (0.0007) 3,102,524	0.594 (0.0007) 3,102,524	0.070 (0.0006) 3,693,228
Sister/Brother	0.035 (0.0005) 5,745,916	0.530 (0.0006) 5,745,916	-0.106 (0.0037) 5,745,916	0.331 (0.0022) 5,745,916	0.035 (0.0005) 6,986,406	0.526 (0.0005) 6,986,406	0.077 (0.0004) 8,181,685

Notes: Except for column 7, each cell in this table is a separate regression of sibling adult lifespan (or of the indicated outcome) on sibling adult lifespan (or on the indicated outcome) including birth cohort fixed effects for each person. We include all unique possible sibling pairs as observations and cluster standard errors 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, the next two columns include all sibling pairs for whom both education and lifespan are available. The final column includes the intergenerational persistence in lifespan between the children in the previous two columns and their fathers. Column 7 is restricted to people that both have a value for education (i.e. can be linked to the 1940 census) and have at least one sibling. In this column we report the father child coefficients instead of the sibling coefficients. This sample is about 13 million individuals. The reason these columns sum to more than 13 million is that there is overlap; sisters of sisters can also be sisters of brothers.

Table 6. Adult Lifespan Coefficients among Siblings and Twins

specification	Siblings		Twins	
	Adult lifespan sibling coefficient	Father's lifespan and child's lifespan	Adult lifespan sibling coefficient	Father's lifespan and child's lifespan
Brother/Brother sample	0.134 (0.0006) 4,125,511	0.084 (0.0005) 4,679,398	0.185 (0.006) 32,679	0.080 (0.0042) 65,358
Sister/Sister sample	0.105 (0.0007) 3,102,524	0.070 (0.0006) 3,693,228	0.162 (0.0064) 29,195	0.070 (0.0043) 58,390
Sister/Brother sample	0.035 (0.0005) 6,986,406	0.077 (0.0004) 8,181,685	0.051 (0.0049) 47,756	0.062 (0.0032) 95,512

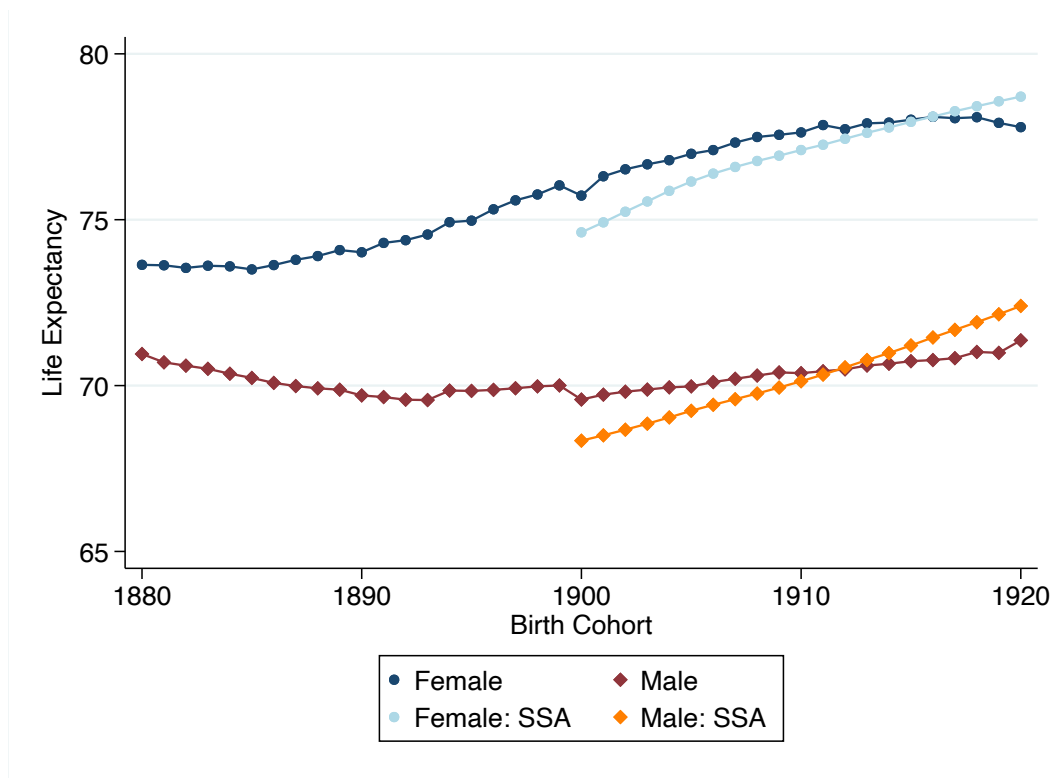
Notes: Each cell in this table is a separate regression. We include all unique possible sibling pairs as observations and cluster standard errors by family. The sample of twins includes all pairs of individuals born in the same year and month within the same family. Columns 1 and 3 are the coefficients when sibling (or twin) adult lifespan is regressed on sibling adult lifespan, including birth cohort fixed effects for each person. Columns 2 and 4 include the intergenerational persistence in lifespan between the children in the previous columns and their fathers. Sample sizes differ across columns because the number of unique sibling pairs in a family often differs from the number of children in a family (i.e. a family with 4 children has 6 sibling pairs).

Table 7. The Contribution of Family Factors to Lifespan and SES in Sibling Samples

	Adult Lifespan	Education
Panel A: Raw sibling correlations		
Adjusted R-squared with just family FE	0.096	0.546
Panel B: Regression of adult lifespan, without family FE		
Adjusted R-squared with just individual covariates	0.039	0.130
Panel C: Regression of adult lifespan, with family FE		
Adjusted R-squared with both	0.131	0.592
N	22,280,230	11,366,597

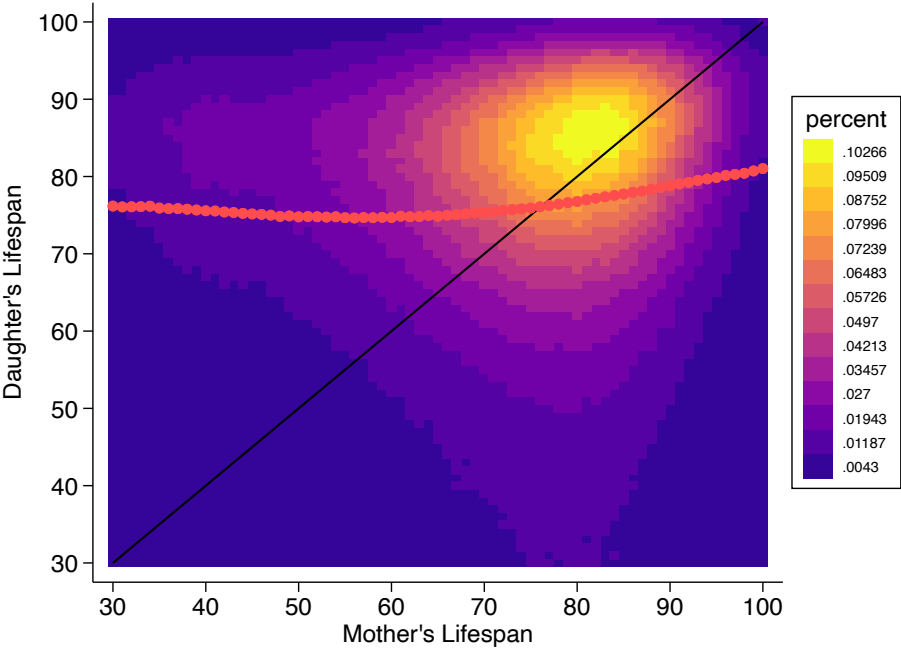
Notes: In this table, we combine all siblings into a single sample. Panel A simply reports the raw sibling correlations in this sample, for reference. Panel B is a regression of the outcome (column header) on covariates: birth cohort of mother FE, birth cohort of father FE, child cohort FE, place of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

Figure 1. Trends in Adult Life Expectancy at Age 25, by Sex

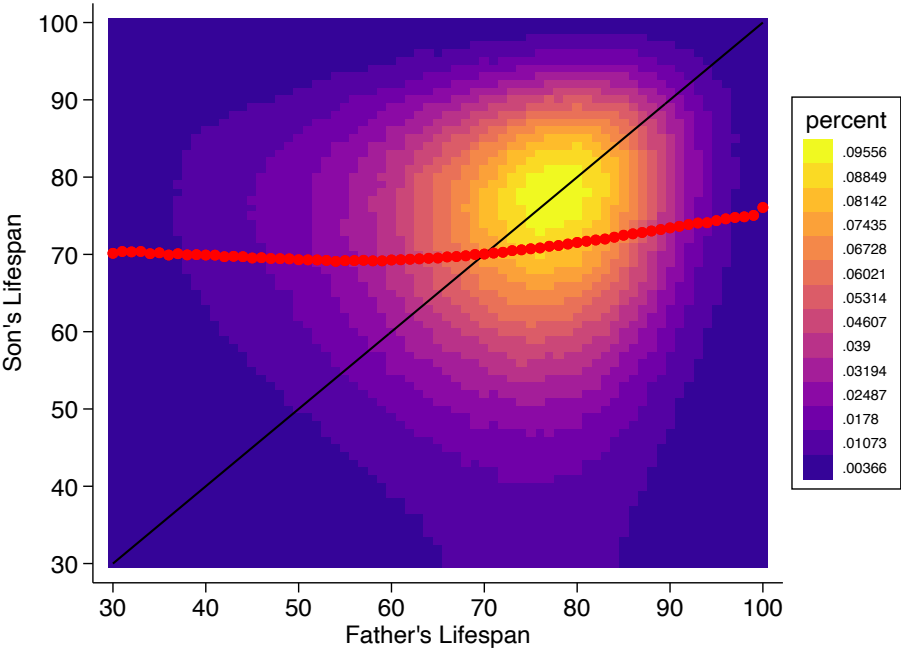


Notes: This figure shows a cohort's adult lifespan (its average/expected age at death conditional on survival to age 25) for cohorts born 1880 to 1920. Points correspond to those who are observed in the 1900-1920 Censuses and in the Census-Family Tree matched data, or are calculated using the Social Security Administration's Cohort Life Tables (<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

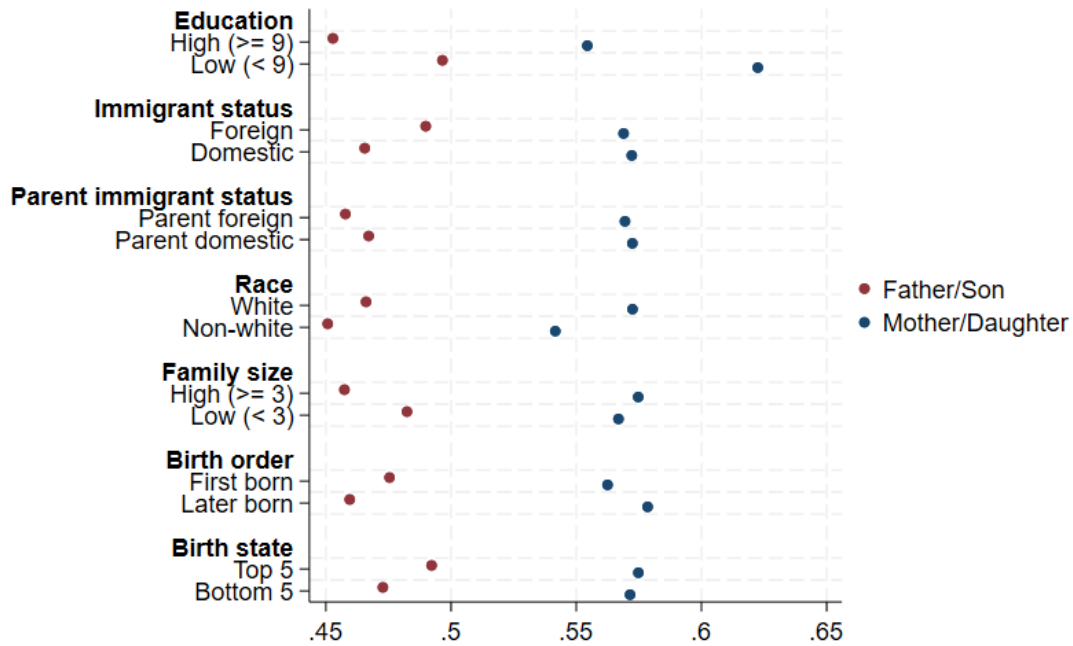


b. Son-Father distribution of age at death

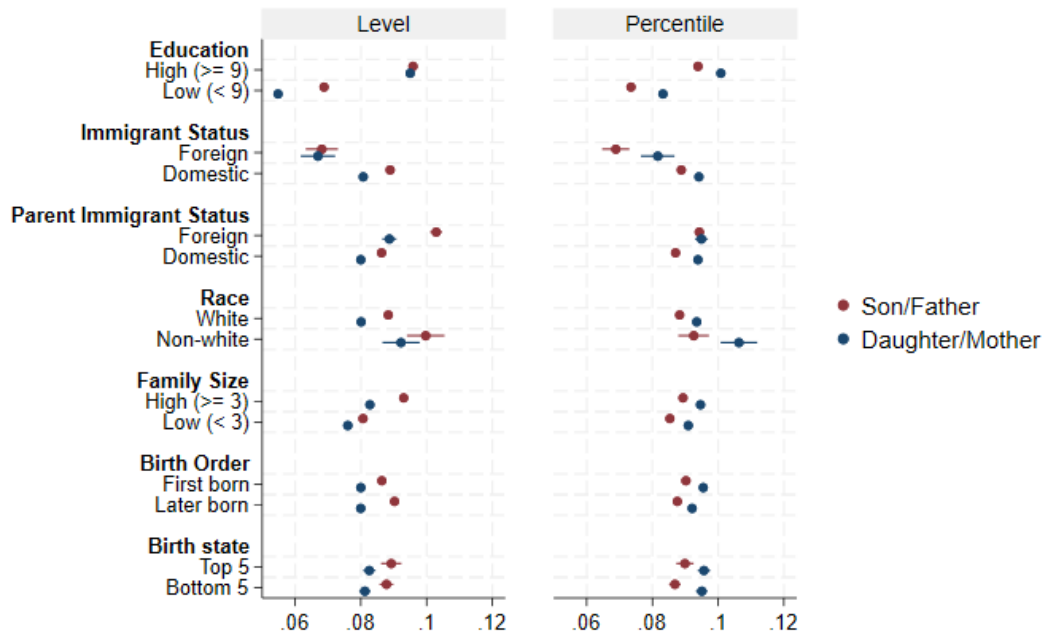


Notes: Panel a shows the joint distribution of the age at death among women born 1880-1920 who survived to age 25, and the age at death of their mothers. Panel b shows the joint distribution for men and their fathers. The black line is a 45-degree line. Each square shows the percent of data falling in that region; for example, 0.1 means that 0.1 percent of the data is in that area. The red points show the average age at death of children for each age at which their parents died.

Figure 3. Heterogeneity in Mobility in Adult Lifespan, by Sex and Group
 a. Absolute mobility

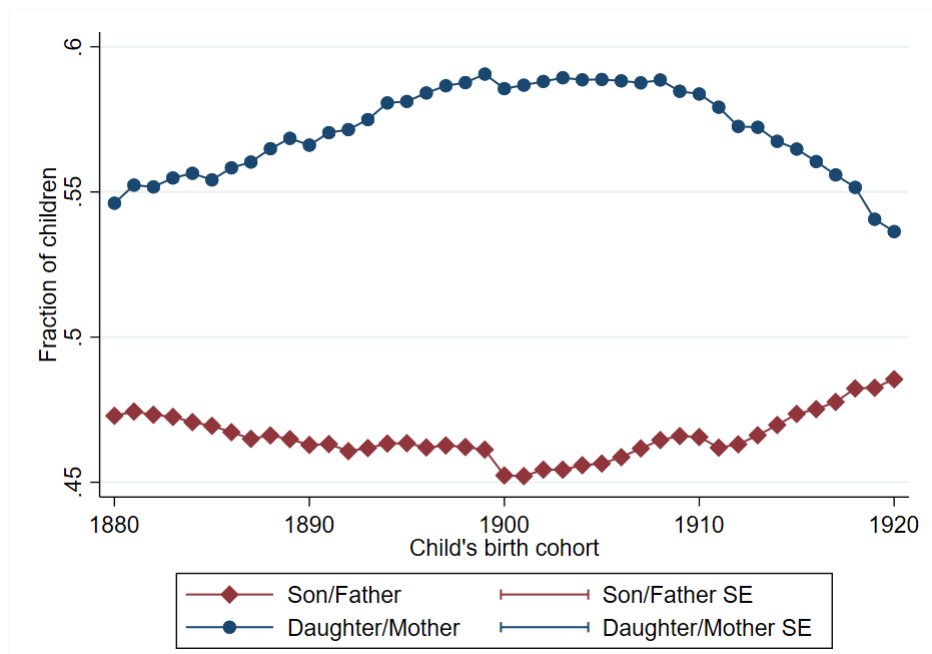


b. Relative persistence



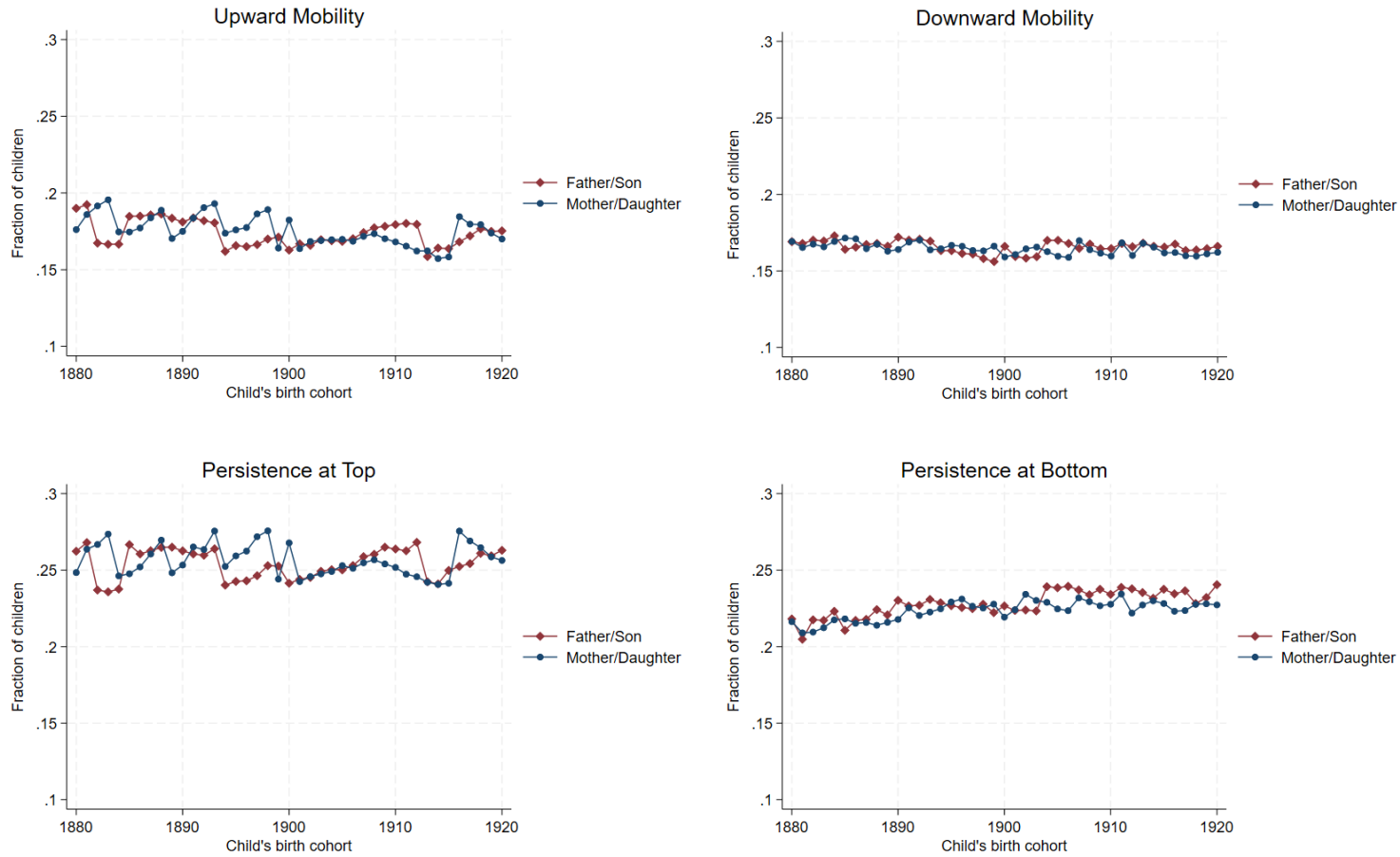
Notes: In Panel a, each circle reports the absolute mobility measure (the probability a child lives longer than their same-sex parent) along with its 95% CI for each of the indicated subsamples, defined by education level, immigrant status, parental immigrant status, family size, birth order or birth state. Panel b reports the intergenerational persistence in lifespan 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. Census Tree data.

Figure 4. Trends in Absolute Mobility, by Sex and Birth Cohort in the US
 Percent of children living longer than same-sex parent



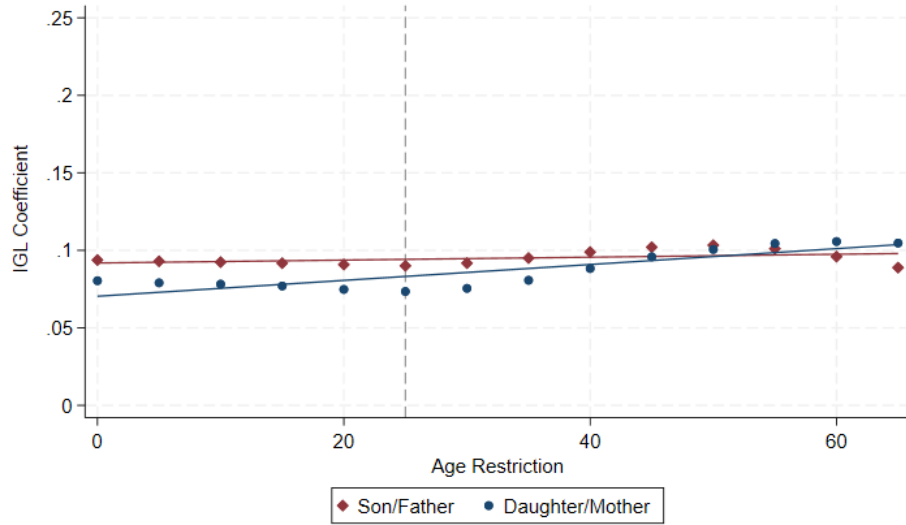
Notes: Each dot is the fraction of children in the sample whose age at death was greater than their parent's age at death (conditional on both parents and children living to age 25). Standard errors are included for each estimate in the figure, but they are too small to see. Census Tree data.

Figure 5. Trends in Relative Mobility in Lifespan



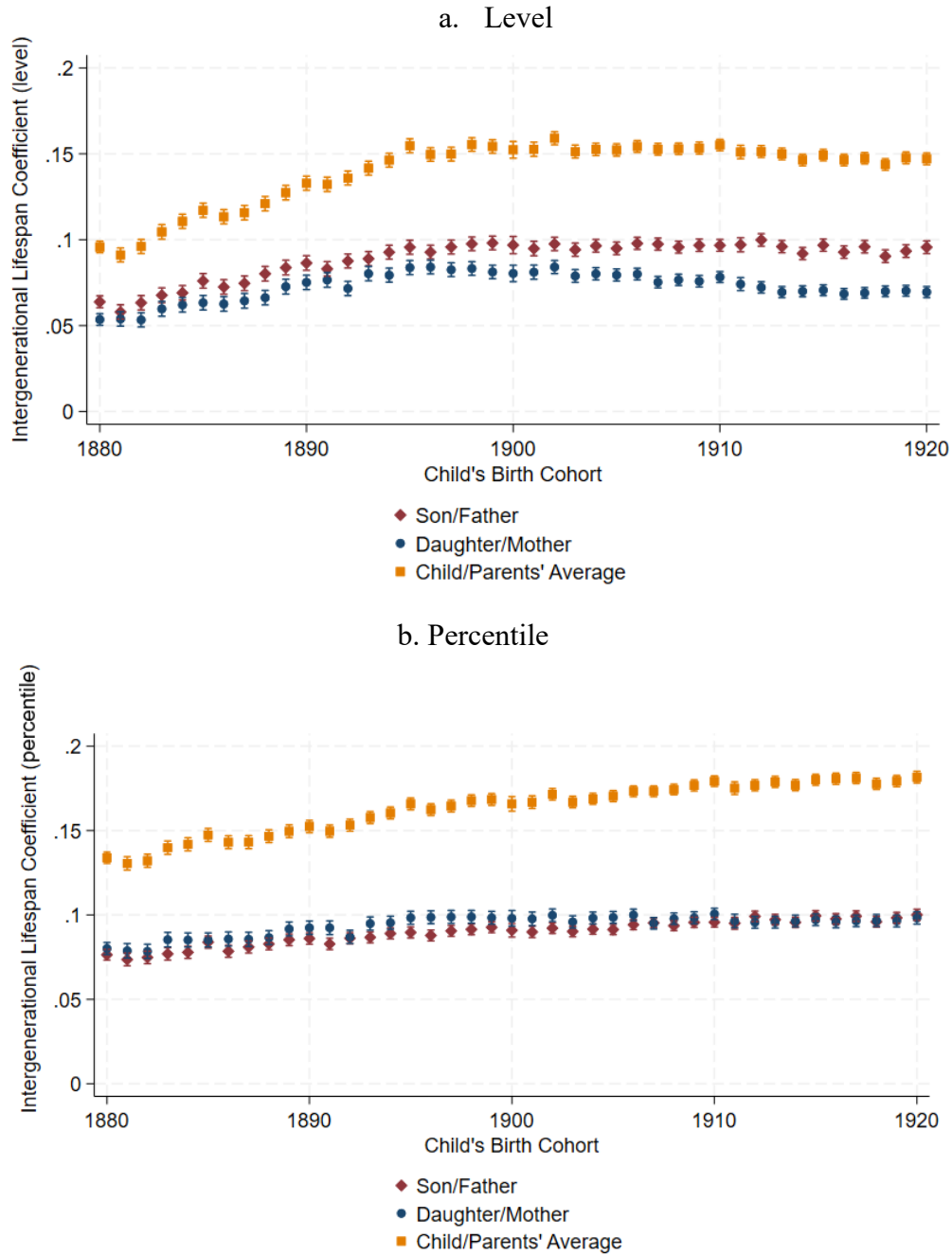
Notes: Upward mobility is the share of children whose parents were at the bottom 20 percent of the sex-specific adult lifespan distribution who end up in the top 20 percent of the sex-specific adult lifespan distribution in their cohort. Downward mobility is the share of children born to parents in the top 20 percentile of the adult lifespan distribution who end up in the bottom 20 percent of the sex-specific adult lifespan 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 lifespan distribution of their cohort. Census Tree data.

Figure 6. Intergenerational Persistence in Lifespan using Different Age Restrictions as Cutoffs



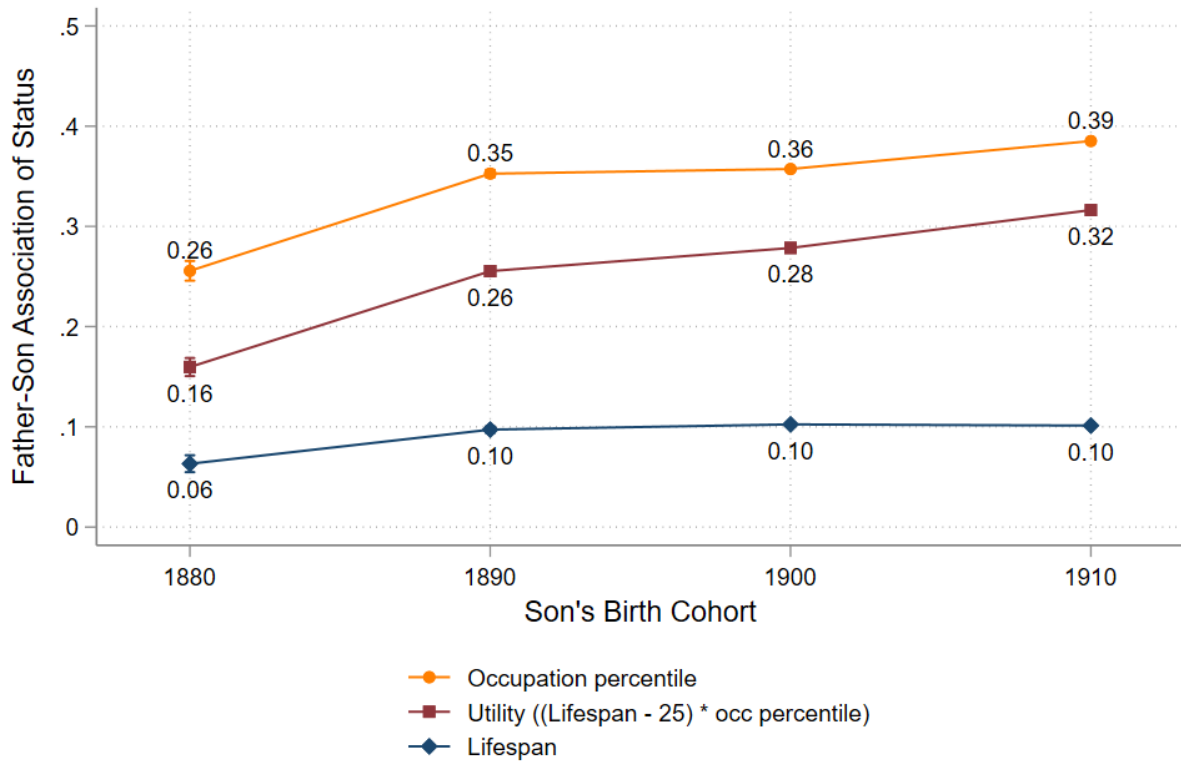
Notes: Each triangle and circle correspond to a coefficient from a regression of the child's lifespan on the parent's lifespan, estimated separately. The specifications in these figures include birth cohort fixed effects for parent and child. The age restriction is applied to both parent and child. For example, if the age restriction on the x-axis is 45, it means we restrict attention to parent-child pairs in which both the parent and the child survive to age 45. Census-Tree data. In the paper we restrict both to survive to age 25.

Figure 7. Changes in the Intergenerational Persistence in Lifespan by Level and Percentile for individuals born 1880-1920



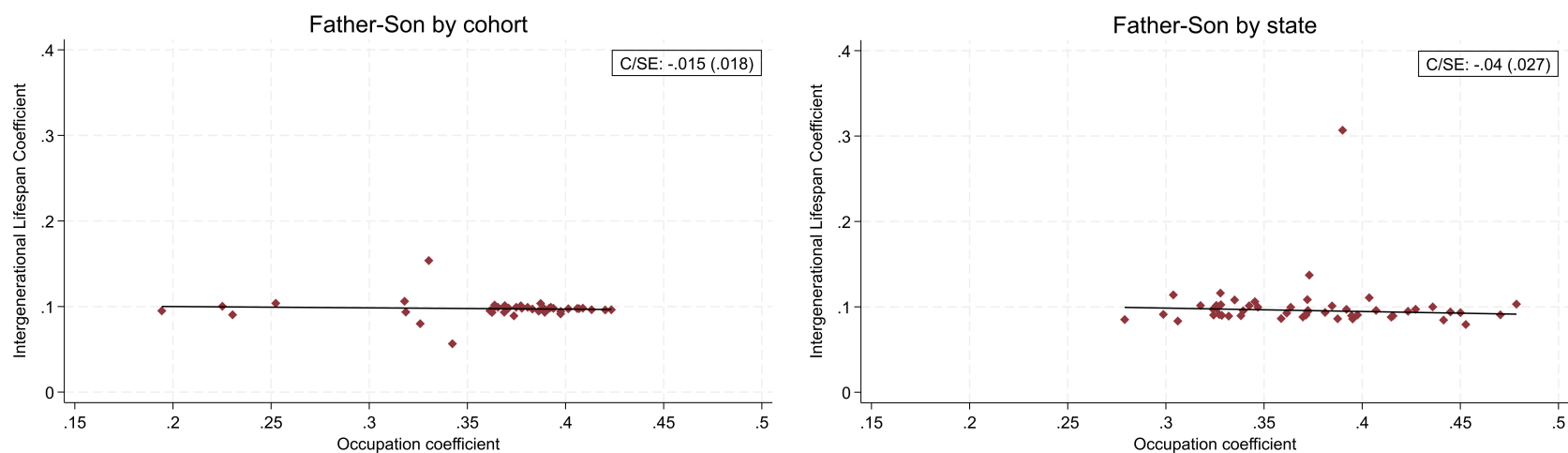
Notes: In this figure, for each birth cohort, we estimate the intergenerational coefficient of lifespan for son/father pairs, daughter/mother pairs, and the coefficients between the child's lifespan and the average lifespan of both parents. Each point corresponds to the regression coefficient of a cohort-specific regression of the child's lifespan on the parent's lifespan controlling for birth cohort fixed effects for the parent. We estimate the regression separately for each birth cohort. Intergenerational lifespan coefficients are estimated using those observations in our dataset in which both parents and children lived to at least age 25.

Figure 8. Intergenerational Correlations in Occupation, Lifespan and Wellbeing



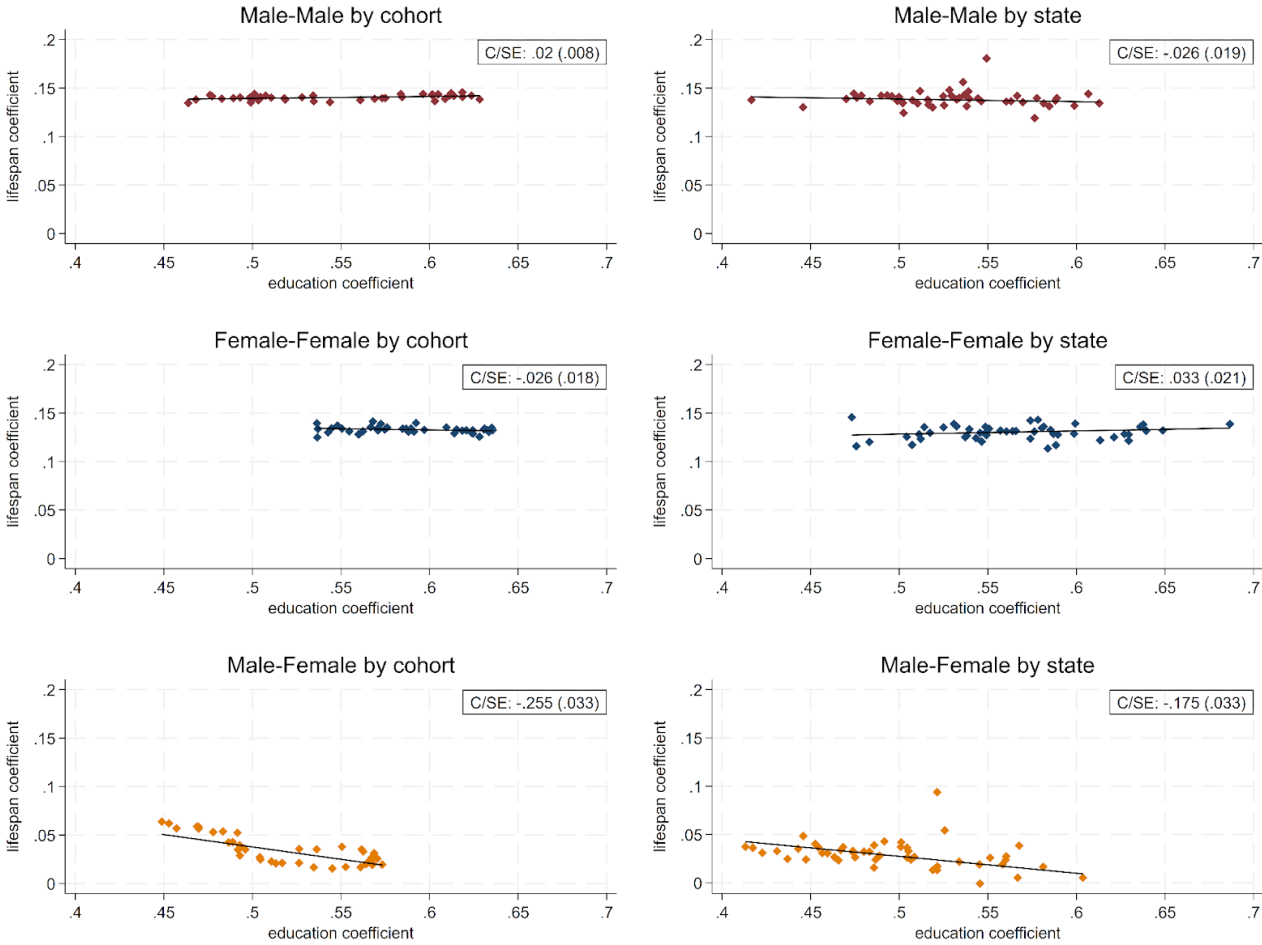
Notes: This figure uses data that is the intersection of our sample with lifespan and the father-son sample from Buckles, Price, and Ward (2023) who report the occupation estimates for census years only. Birth cohorts are binned by decade (i.e. 1880-1889 is the 1880 cohort). Standard errors are present, but too precise to be legible outside of the 1880 cohort.

Figure 9: Relationship between the Intergenerational Associations in Adult Lifespan and Intergenerational Associations in Occupation-based Income Percentiles among Men born 1880-1920



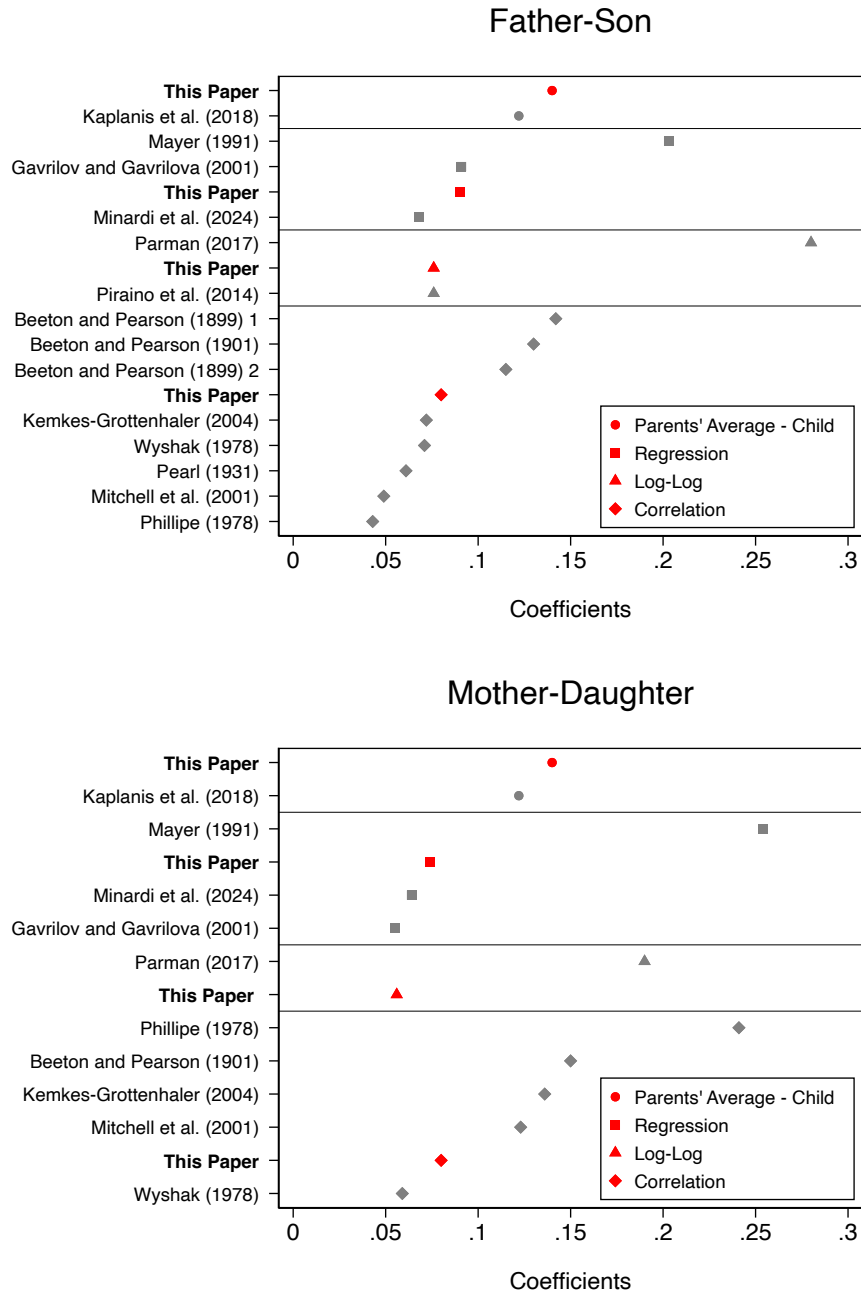
Notes: These figures plot the regression coefficients in lifespan on the y-axis against the coefficient in occupation-based percentile on the x-axis for a given cohort (left panel) or birth state (right panel). We use the (inverse of the) variance of the occupation-coefficient as weights in these regressions, though the unweighted results are very similar. The sample is restricted to men and fathers who have data on both and who survived to age 25. The line is the best linear fit, and the estimated slope and its standard error are given in the box in top right of each panel.

Figure 10. 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 lifespan on the y-axis against the sibling regression coefficients in education on the x-axis for a given cohort or state. Fitted lines are weighted by the in-sample population of the relevant state or cohort. We use the (inverse of the) variance of the education coefficient as weights in these regressions and figures, though the unweighted results are very similar. We restrict attention to individuals surviving to age 60.

Figure 11. Comparison to Previous Estimates of Intergenerational Persistence in Lifespan



Notes: The figure reports the estimates from various publications. The estimates from this paper come from Table 4 and Appendix Table 3. 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. At the end of Table A.1, we specify the exact location of each estimate in the original publication.

Appendix Tables and Figures

Table A.1. Previous Estimates of the Intergenerational Correlations in Lifespan

Paper	Correlation 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

Minardi et al. (2024)	Father-Son: 0.068 Father-Daughter: 0.067 Mother-Son: 0.054 Mother-Daughter: 0.064	Father-Son: 0.003 Father-Daughter: 0.004 Mother-Son: 0.002 Mother-Daughter: 0.003	Father-Son: 208,784 pairs Father-Daughter: 142,100 pairs Mother-Son: 208,784 pairs Mother-Daughter: 142,100 pairs	US	1690-1910
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		122,766	Cape Colony, South Africa	1652 - 1850

	conditioned on survival post 15) Sibling-Sibling: 0.171 (0.086 if conditioned on survival post 15)				
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	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	Male-male MZ twin pairs: 1567 (reared together), 82 (reared apart) Male-male DZ twin pairs: 2814 (reared	Swedish Twin Pairs	1886-1925

	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)	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)	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)		
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

We use the following estimates from these papers to create Figure 11 in the paper.

- 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 Mecklenburg 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 Phillippe (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.
- Father-Son (Mother-Daughter) correlation for Minardi et al. (2024) is from Table A.5 (Table A.5), which uses 208,784 (142,100) pairs drawn from cohorts born in the United States between 1690 and 1910, with the age restriction of surviving until 30.
- 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 lifespan and they do not report the correlations for Mother-Daughter and Father-Son pairs. The data come from [Geni.com](https://www.geni.com) where individual users can upload family tree information.

Table A.2. Comparing Tree data with SSA data by cohort

Age	Sample		SSA		Difference	
	Male	Female	Male	Female	Male	Female
1900 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
1910 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 at different ages. Difference calculated (CensusTree sample-SSA), 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	1900	Matched	1920
	Analysis	Census	Analysis	Census
	Data	Data	Data	Data
	1880-1899	1880-1899	1900-1920	1900-1920
	cohorts	cohorts	cohorts	cohorts
Lifespan	72.10		73.68	
Female	0.47	0.50	0.47	0.50
Birth Year	1890	1890	1910	1910
White	0.99	0.86	0.99	0.88
Black	0.01	0.13	0.01	0.09
Northeast	0.15	0.22	0.14	0.24
Midwest	0.43	0.33	0.38	0.29
South	0.32	0.35	0.37	0.35
West	0.05	0.04	0.08	0.06
Immigrant	0.02	0.12	0.01	0.02
Father is Immigrant	0.16	0.21	0.11	0.20
Mother is Immigrant	0.12	0.16	0.09	0.16
Observations	11,555,193	33,567,608	14,578,968	45,560,832

Notes: The estimates in this table compare the mean values of individuals in either the 1900 or 1920 census to individuals from those censuses who match to information on their lifespan and the lifespans of both of their parents.

Table A.4. Raw Correlations in Lifespan Across Generations

Model	Outcome			Observations
	Lifespan (Years)	Percentile	Log Lifespan	
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. Assessing the Impact of WWI, WWII and the 1918 Flu Pandemic

Model	Outcome: Lifespan (Years)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	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.	(4) + Without Key Death Years	# of Obs.
Son/Father	0.089 (0.0003)	0.090 (0.0003)	0.087 (0.0003)	0.087 (0.0003)	13,944,386	0.088 (0.0004)	11,295,899
Son/Mother	0.062 (0.0003)	0.062 (0.0003)	0.059 (0.0003)	0.059 (0.0003)	13,944,386	0.060 (0.0003)	11,295,899
Son/Parents' Average	0.140 (0.0004)	0.141 (0.0004)	0.137 (0.0004)	0.137 (0.0004)	13,944,386	0.138 (0.0005)	11,295,899
Daughter/Father	0.075 (0.0004)	0.075 (0.0004)	0.072 (0.0004)	0.072 (0.0004)	12,189,775	0.072 (0.0004)	9,956,998
Daughter/Mother	0.081 (0.0003)	0.074 (0.0003)	0.071 (0.0003)	0.071 (0.0003)	12,189,775	0.071 (0.0003)	9,956,998
Daughter/Parents' Average	0.150 (0.0005)	0.142 (0.0005)	0.138 (0.0005)	0.138 (0.0005)	12,189,775	0.137 (0.0005)	9,956,998

Notes: Each cell separately provides the estimated regression coefficient in lifespan between the two individuals indicated in the row label. Standard errors are clustered at the family level. 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 (6) drops observations if the mother, father, or child died in one of the following years: 1918, 1919, 1920, 1942, 1943, and 1944, which is done to account for deaths from the World Wars and the Spanish flu pandemic.

Table A.6. Sensitivity to Alternative Censoring Rules

	Both individuals live to age 25	Children live to age of parent at child's birth
Son/Father	0.0901 (0.0003) 13,944,386	0.0868 (0.0003) 13,481,225
Daughter/Mother	0.0735 (0.0003) 12,189,775	0.0698 (0.0003) 12,033,201

Notes: The specifications in these figures include birth cohort fixed effects for parent and child. These estimates correspond to column 2 of table 3.

Table A.7. Summary Statistics of Sibling Subsample

	Census based sample matched to FamilySearch, cohorts born 1880-1920	
	Full Sample	Siblings
Average Lifespan	72.97 (16.09)	75.601 (13.65)
Father's Lifespan	71.66 (13.56)	72.25 (13.24)
Mother's Lifespan	72.31 (15.89)	72.84 (15.50)
Birth Year	1901	1901
White	0.99	0.99
Black	0.01	0.004
<u>Place of birth and ancestry</u>		
Northeast	0.15	0.14
Midwest	0.41	0.42
South	0.35	0.35
West	0.07	0.08
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 (2.36)	3.42 (2.21)
Birth order	2.39 (1.68)	2.65 (1.72)
Age mother at birth	33.93 (8.02)	34.01 (7.83)
Age father at birth	29.13 (6.71)	29.21 (6.58)
Observations	26,134,160	13,107,678

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 deviations in parentheses.

Table A.8. Sibling correlations for outcomes in the 1940 census

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Individuals with data on lifespan, education, individual income and household income				Individuals with data on lifespan and education		Individuals with data on lifespan
Model	Adult Lifespan	Education	Income	HH Income	Adult Lifespan	Education	Father-child lifespan
Brother/Brother	0.141 3,664,460	0.553 3,664,460	0.260 3,664,460	0.347 3,664,460	0.141 4,126,499	0.551 4,126,499	0.094 4,680,402
Sister/Sister	0.106 2,402,338	0.593 2,402,338	0.167 2,402,338	0.359 2,402,338	0.106 3,102,766	0.585 3,102,766	0.098 3,693,559
Sister/Brother	0.037 5,747,644	0.529 5,747,644	-0.104 5,747,644	0.328 5,747,644	0.037 6,988,569	0.525 6,988,569	0.094 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 intergenerational persistence in lifespan between the children in the previous two columns and their fathers.

Table A.9. Cross-Domain Associations in Lifespan and SES for Sons and Fathers

Outcome:	Son's LS – Father's LS	Son's Occ rank – Father's Occ rank
Panel A: Means	-1.508 [17.919]	-2.483 [29.559]
Outcome:	Son's LS – Father's LS and Son's Occ rank – Father's Occ rank	
Measure :	Correlation	Regression Coefficient
Panel B: Associations	0.0213 4,855,286	-0.0090 (0.0029) 4,855,286

Notes: This table uses data that is the intersection of our sample with lifespan and the father-son sample from Buckles et al. (2023) who report occupation estimates. Panel B column 2 includes birth year fixed effects for both sons and fathers, with errors clustered by family.

Table A.10. Cross-Domain Associations in Lifespan and SES for Siblings

	LS – Sib LS	Inc – Sib Inc
Panel A: Means	0.042 [18.188]	3.087 [1075.714]
	LS – Sib LS and Inc – Sib INC	
	Correlation	Regression
Panel B: Associations		
Brother/Brother	0.025 3,664,460	0.0004 (0.0001) 3,664,460
Sister/Sister	0.009 2,402,338	0.0003 (0.0001) 2,402,338
Sister/Brother	-0.166 5,747,644	-0.0027 (0.0001) 5,747,644

Notes: Panel B column 2 includes birth year fixed effects for both siblings, with errors clustered by family.

Table A.11. Lifetime utility computations among men in the sibling sample

	(1)	(2)	(3)	(4)	(5)
Outcome	Adult lifespan	Occupation percentile	Annual HH Income	Wellbeing $u(y)$ $= LE_{25}$ $* pct(y)$	Wellbeing $u(y)$ $= LE_{25}$ $* \sqrt{hh inc}$
Brother/ Brother	0.1372 (0.001) 1,194,765	0.3263 (0.001) 1,194,765	0.3289 (0.0036) 1,194,765	0.4247 (0.0012) 1,194,765	0.4663 (0.0013) 1,194,765
Father/ Son	0.0979 (0.001) 1,570,255	0.3504 (0.0012) 1,570,255		0.281 (0.0013) 1,570,255	

Notes: Each cell reports the coefficient of regression of a child's outcome on their sibling's outcome or of a child's outcome on their father's outcome. There is one observation per sibling pair. The sample includes only siblings without missing data in the 1940 census on all outcomes and who survived to age 25. The wellbeing measures in columns 4 and 5 are computed as specified in the column where "pct(y)" refers to the occupation percentile (used in column 2) and "hh inc" refers to household income (used in column 3). This sample is smaller than our sibling samples in Table 5 because we only include individuals who have non-missing data on occupation and whose fathers also have occupation. In Table 5 we only restrict to parents having lifespans.

Table A.12. 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 (0.0004)	0.079 (0.0004)	0.078 (0.0004)	0.078 (0.0004)	0.055 (0.0004)	0.052 (0.0004)	0.052 (0.0004)	0.052 (0.0004)
Child's Education		0.246 (0.002)	0.248 (0.002)	0.267 (0.002)		0.241 (0.002)	0.246 (0.002)	0.267 (0.002)
Income/100			-0.001 (0.001)	0.006 (0.01)			-0.004 (0.001)	0.003 (0.001)
Occupation				-0.020 (0.001)				-0.022 (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 (0.0004)	0.064 (0.0004)	0.064 (0.001)	0.064 (0.001)	0.064 (0.0004)	0.058 (0.0004)	0.057 (0.0004)	0.057 (0.0004)
Child's Education		0.391 (0.002)	0.383 (0.002)	0.382 (0.002)		0.374 (0.002)	0.369 (0.002)	0.369 (0.002)
Income/100			0.017 (0.002)	0.011 (0.002)			0.010 (0.002)	0.007 (0.002)
Occupation				0.004 (0.001)				0.002 (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.13. Assessing how the quality of the age at death information affects the results.
Lifespan sibling coefficient.

Sample:	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 (0.001)	0.069 (0.001)	0.118 (0.002)	0.081 (0.002)
N	2,402,338	3,693,559	229,196	321,367
Panel B: Brother/Brother	0.134 (0.001)	0.084 (0.001)	0.159 (0.002)	0.096 (0.001)
N	3,664,460	4,680,402	542,232	702,565
Panel C: Sister/Brother	0.035 (0.001)	0.077 (0.001)	0.060 (0.001)	0.091 (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

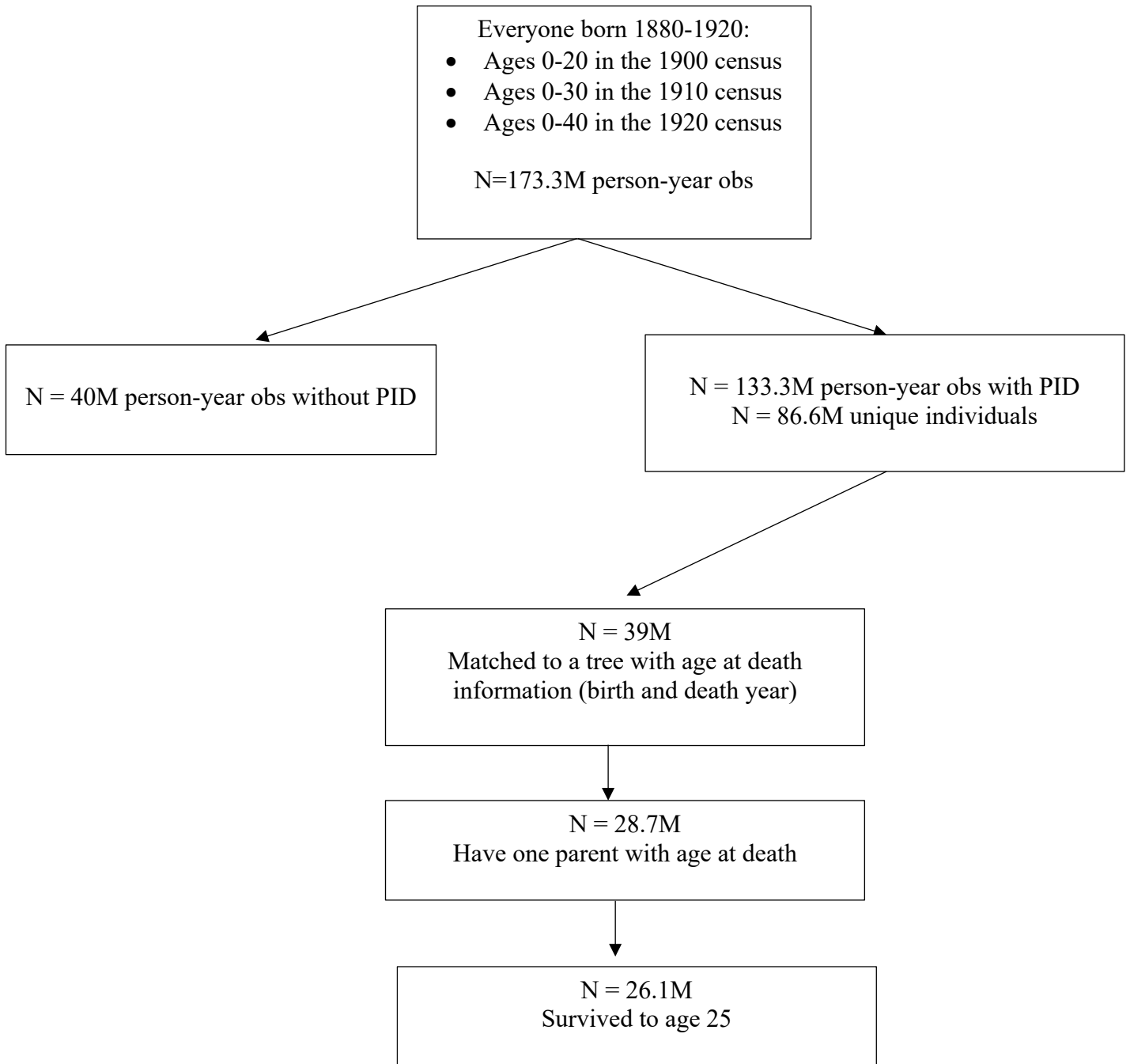
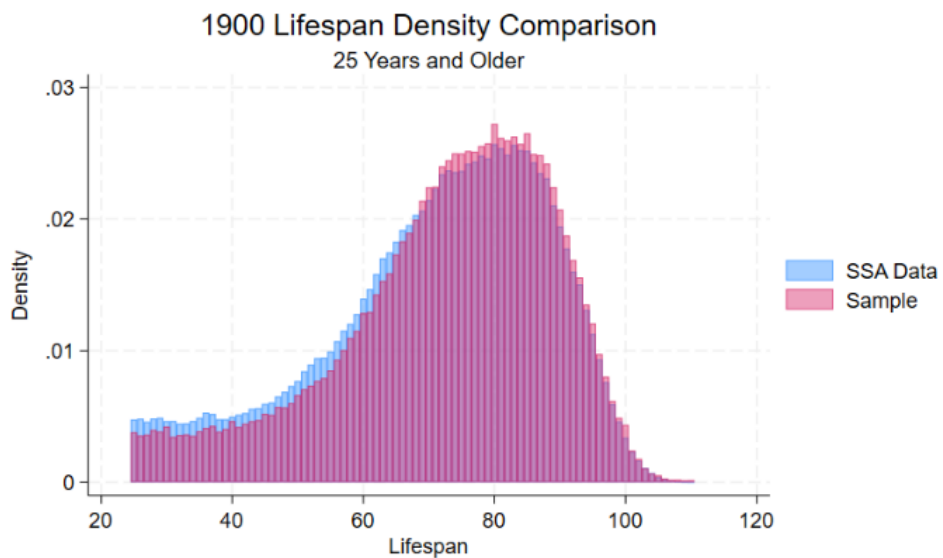
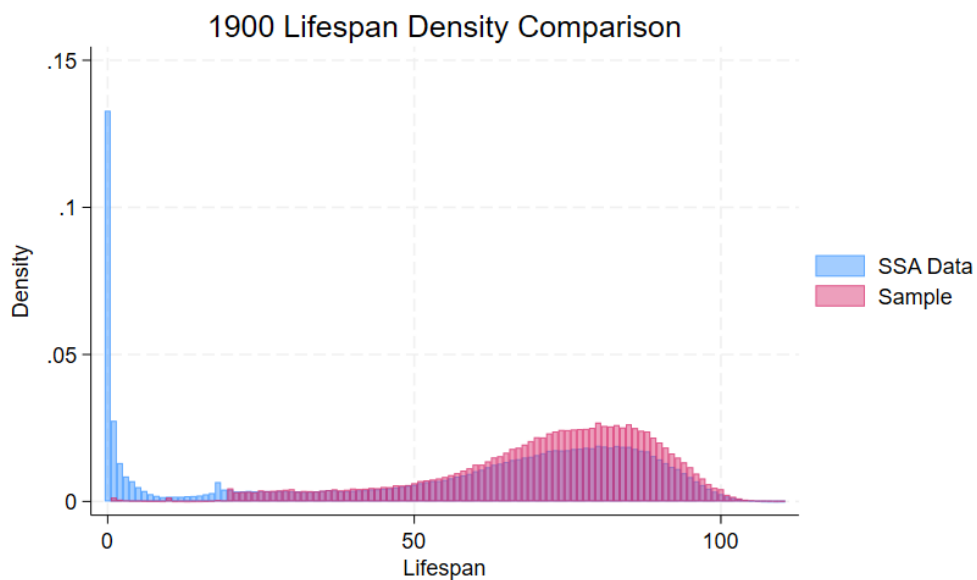
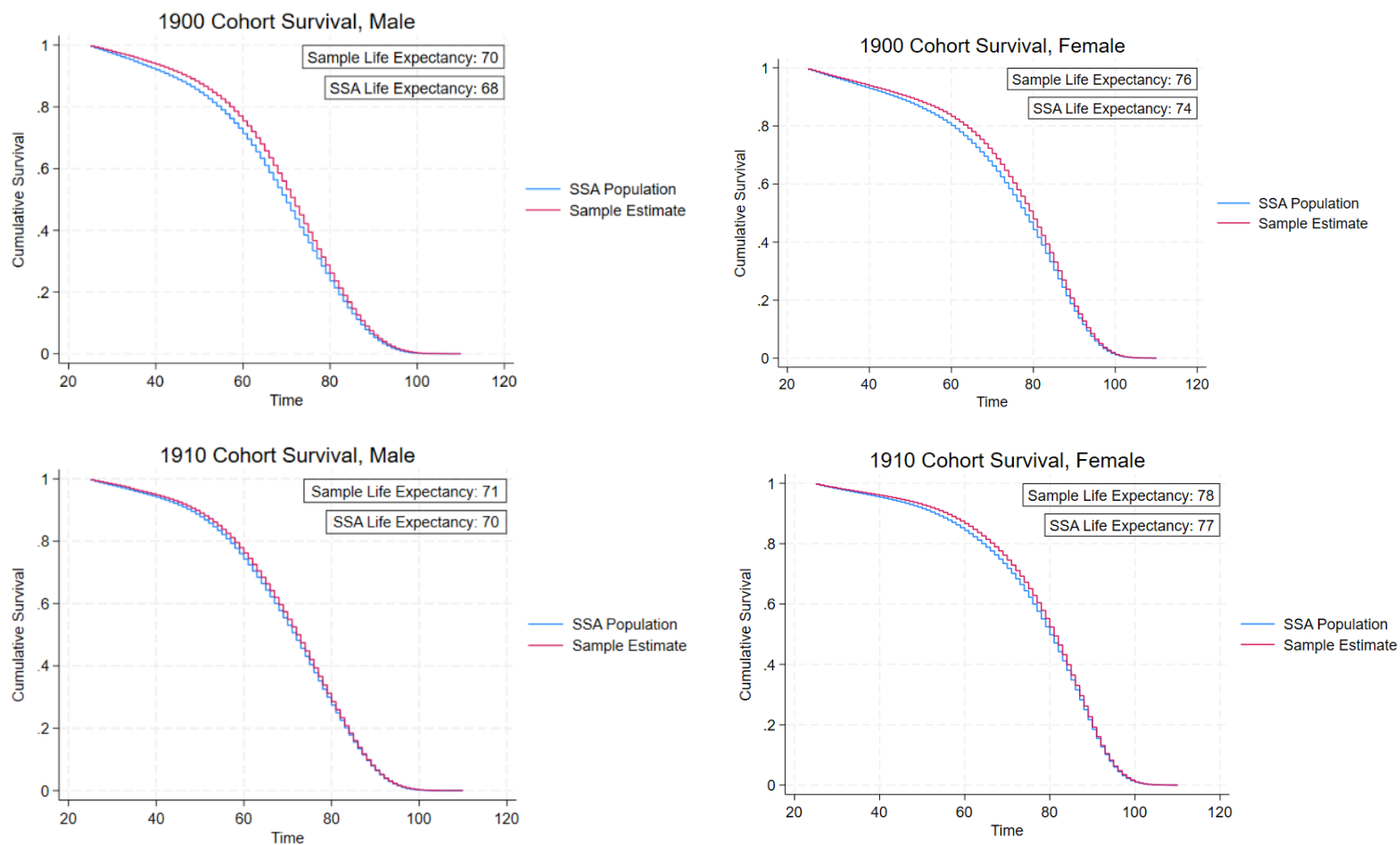


Figure A.1. Distribution of the age at death In the Census-Tree data
Comparison with SSA Cohort Data



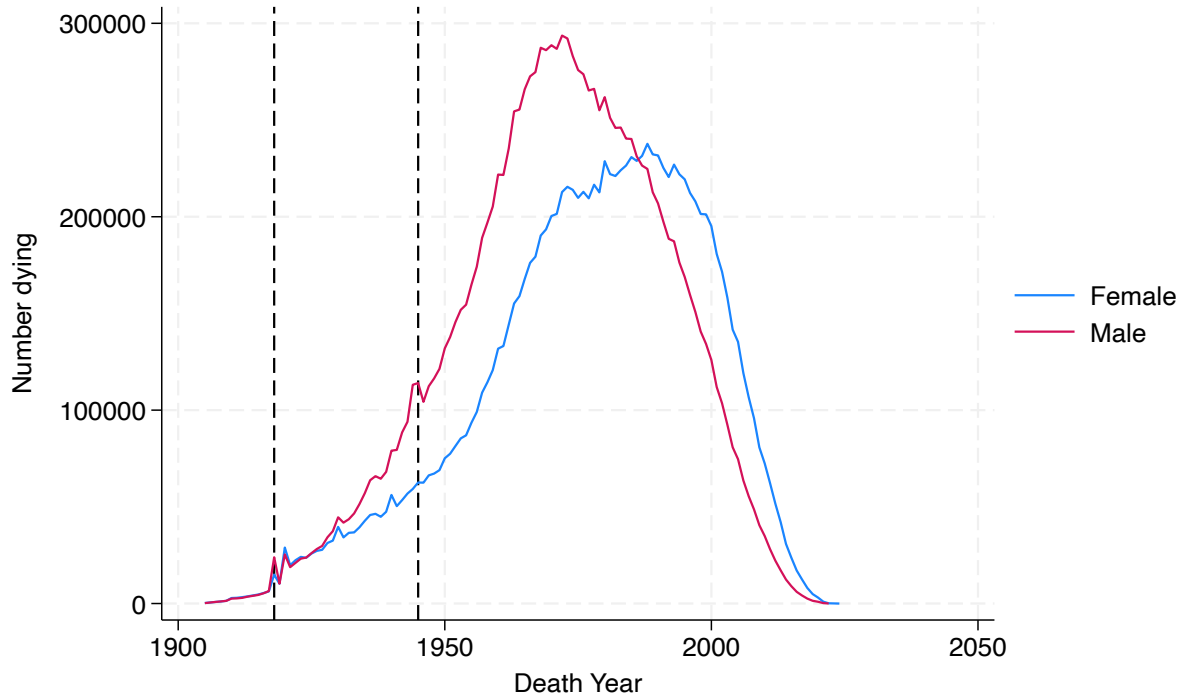
Notes: These figures use our sample (derived from the FamilyTree, see text for details) and cohort life tables produced by the Social Security Administration (SSA), available here: https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf.

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



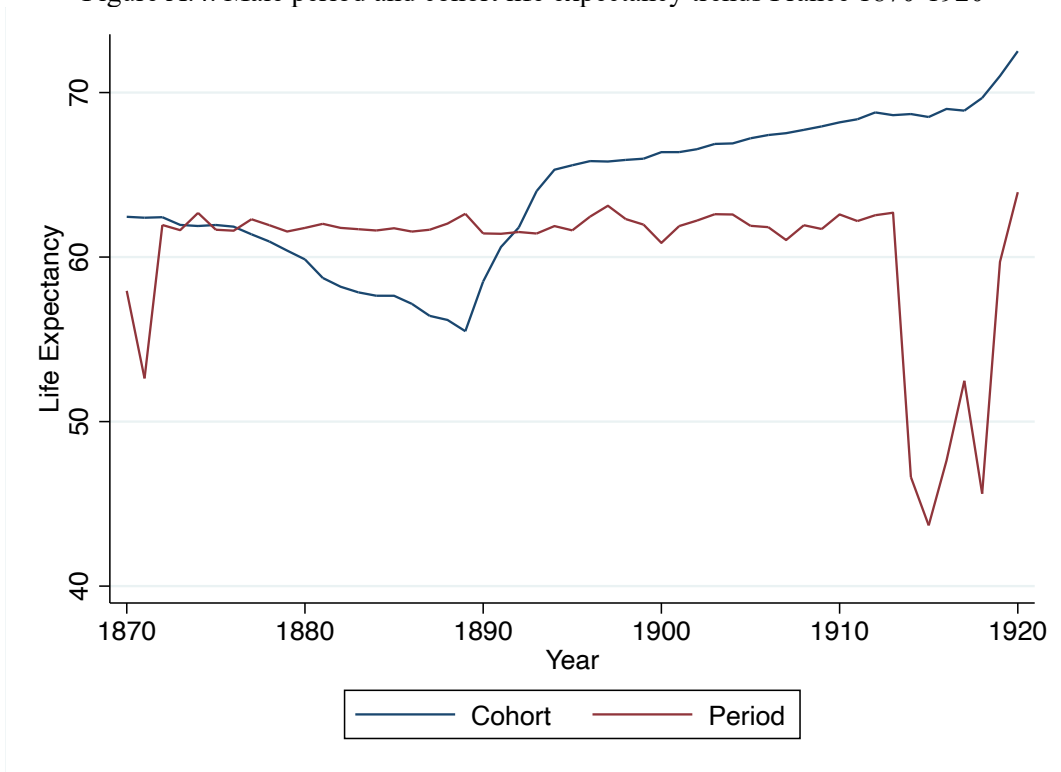
Notes: These figures use our sample (derived from the Family Tree, see text for details) and cohort life tables produced by the Social Security Administration (SSA), available here: https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf. Kaplan-Meier estimates are produced using the methods described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



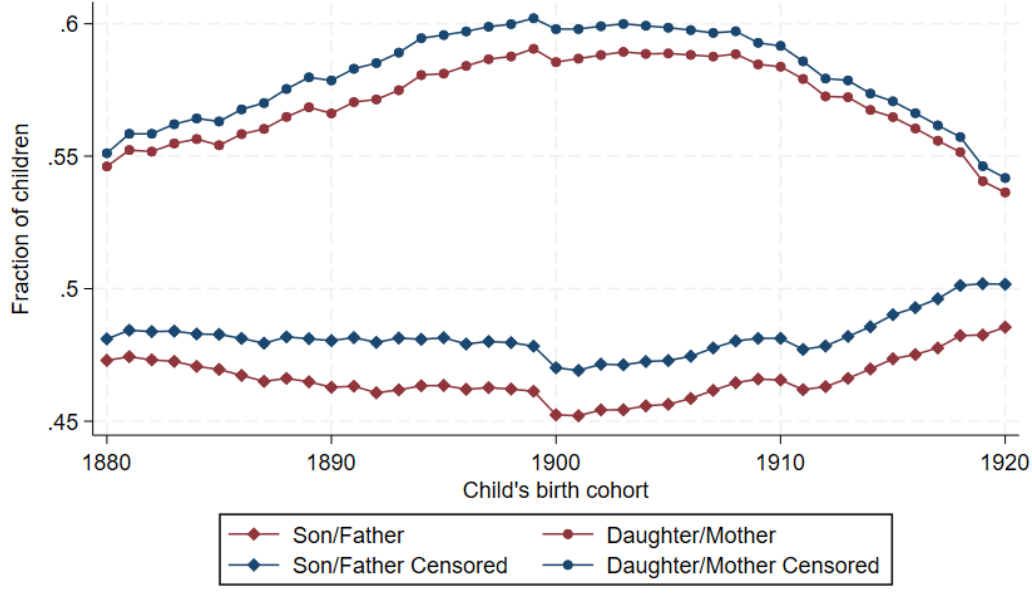
Notes: The figure shows the death year of individuals born 1880-1920 who survived to age 25 and who are in our sample (they have birth and death dates and so do their parents). The dashed lines denote 1918 and 1945, the deadliest years of WWI and WWII. In 1918 there was also a flu pandemic.

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



Notes: This figure uses period and cohort tables from the Human Mortality Database.

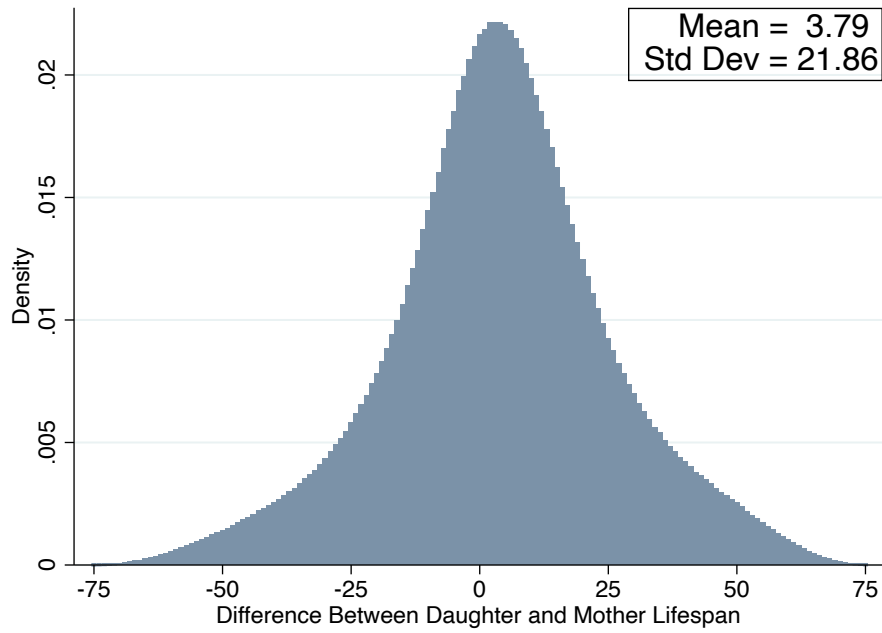
Figure A.5. Trends in absolute mobility, by sex and birth cohort
 Percent of children living longer than same-sex parent using alternative censoring rules



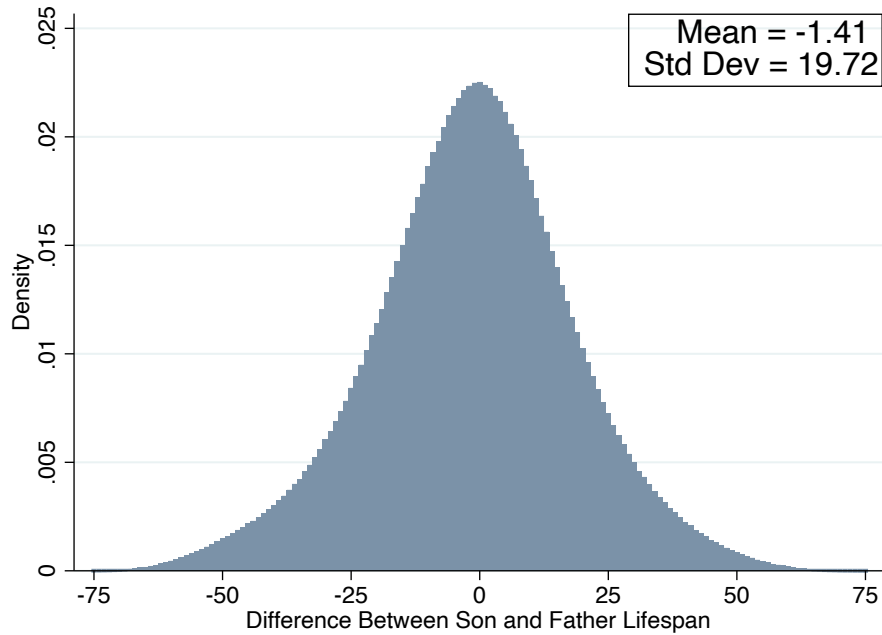
Notes: Each dot is the fraction of children in the sample whose age at death was greater than their parent's age at death. The blue lines show estimates in which both parent and child are required to survive to age 25. The red lines show the estimates for the alternative censoring rule in which children are required to survive to the same age as their parents at the time of their birth. Standard errors are included for each estimate in the figure, but are too small to see.

Figure A.6. Distribution of adult lifespan gap in Census-tree data

a. Daughter age at death – mother age at death

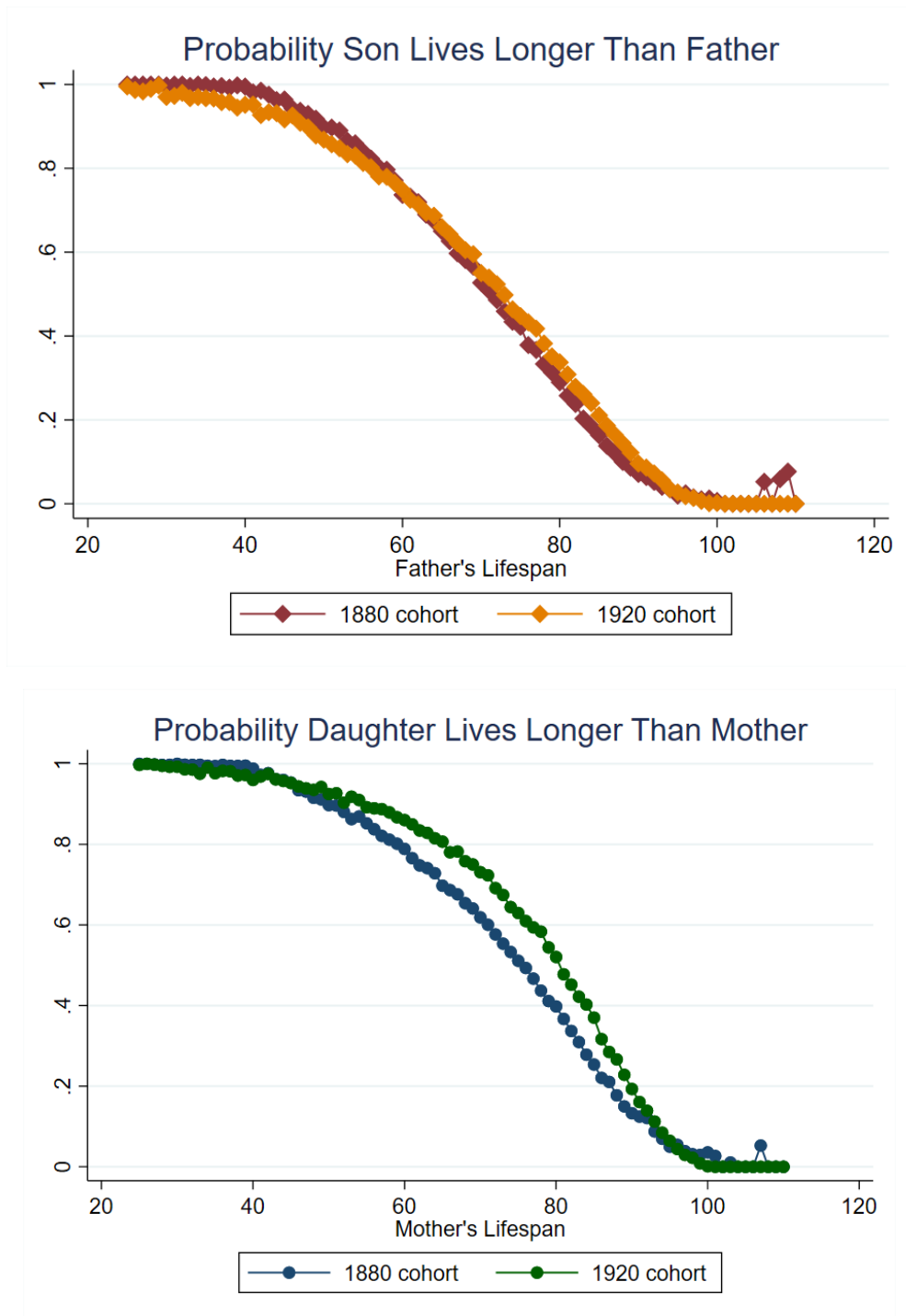


b. Son age at death – father age at death



Notes: The figure shows the 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.

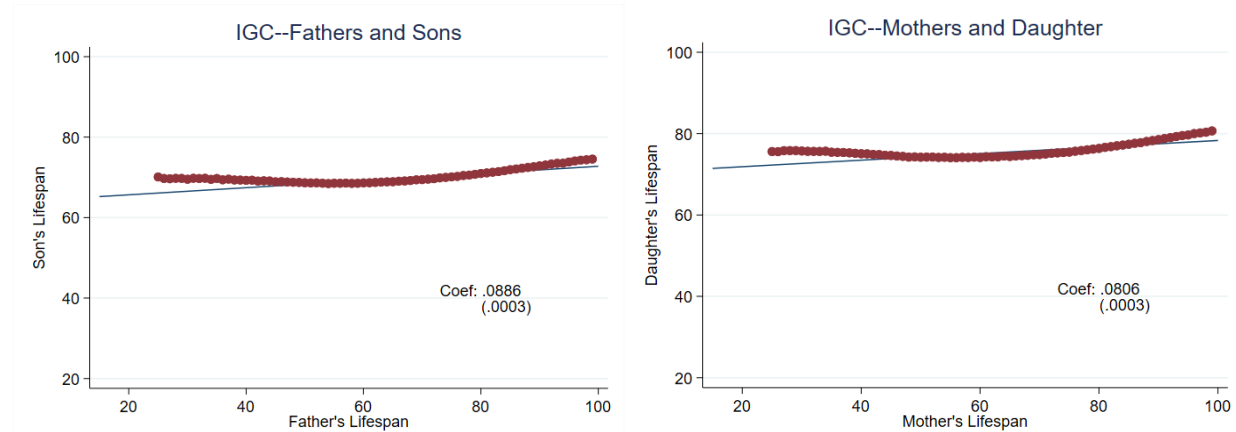
Figure A.7. Changes in Absolute Mobility 1880-1920



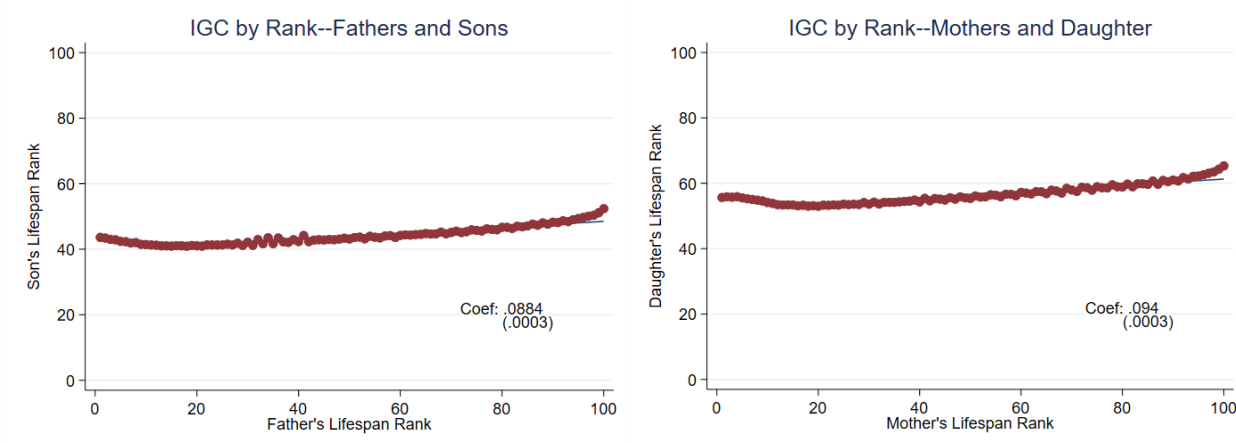
Notes: The figures use the Census-Tree data for cohorts born 1880-1920. They show the survival curves for parents and children. The top panel reports the curves for fathers and sons, the bottom panel reports the curves for mothers and daughters.

Figure A.8. Test for Linearity of the child-parent lifespan relationship, by sex

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



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



Notes: The top panels 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 panels 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. The thin blue line shows the best linear prediction; the estimated slope and its standard error for each figure is reported under "coef" inside each figure.

Appendix A

1 Covariances from Simple Model of Lifespan

For the model of lifespan described in the paper, this appendix outlines the covariances in lifespan between parents, parents and children, siblings, twins, and grandparents and grandchildren.

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

$$\begin{aligned} Cov(L_{m(i)}, L_{f(i)}) = & Cov(\delta G_{m(i)} + \alpha_{r(m(i))c(m(i))} + \gamma_{m(i)}^{s(m(i))} + \lambda SES_{m(i)} + \varepsilon_{m(i)}, \\ & \delta G_{f(i)} + \alpha_{r(f(i))c(f(i))} + \gamma_{f(i)}^{s(f(i))} + \lambda SES_{f(i)} + \varepsilon_{f(i)}) \end{aligned}$$

$$Cov(L_{m(i)}, L_{f(i)}) = \delta^2 Cov(G_{m(i)}, G_{f(i)}) + Cov(\alpha_{r(m(i))c(m(i))}, \alpha_{r(f(i))c(f(i))}) + \lambda^2 Cov(SES_{m(i)}, SES_{f(i)})$$

where the first term denotes the extent to which there is assortative mating between spouses based on genes. The second term is the covariance of the environmental component. The third term denotes the extent to which there is assortative mating between spouses based on socioeconomic status.

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

$$\begin{aligned} Cov(L_{f(i)}, L_i) = & Cov(\delta G_{f(i)} + \alpha_{r(f(i))c(f(i))} + \gamma_{f(i)}^{s(f(i))} + \lambda SES_{f(i)} + \varepsilon_{f(i)}, \\ & \delta \frac{1}{2} (G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i) \end{aligned}$$

As discussed in the paper, the environmental component ($\alpha_{r(i)c(i)}$), the within-family sex-specific random effect ($\gamma_i^{s(i)}$), and SES (SES_i) are not correlated with the genes (G_i), individual specific shocks (ε_i), and each other. As such, the covariance between father and son simplifies to the following and is reported as Equation (2) in the paper:

$$\begin{aligned} Cov(L_{f(i)}, L_i) = & \delta^2 \frac{1}{2} Cov(G_{f(i)}, G_{m(i)}) + \delta^2 \frac{1}{2} V(G_{f(i)}) + Cov(\alpha_{r(f(i))c(f(i))}, \alpha_{r(i)c(i)}) \\ & + Cov(\gamma_{f(i)}^1, \gamma_i^1) + \lambda^2 Cov(SES_{f(i)}, SES_i) \end{aligned}$$

The covariance between father and son depends on the genetic assortativeness of the parents, the extent to which father and son share genes, the extent to which they share an environment, the covariance of the within-family sex-specific random effect, and the covariance in the socioeconomic status component between fathers and sons.

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

$$\begin{aligned} Cov(L_{m(i)}, L_i) &= Cov(\delta G_{m(i)} + \alpha_{r(m(i))c(m(i))} + \gamma_{m(i)}^{s(m(i))} + \lambda SES_{m(i)} + \varepsilon_{m(i)}, \\ &\quad \delta \frac{1}{2}(G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i) \\ Cov(L_{m(i)}, L_i) &= \delta^2 \frac{1}{2} Cov(G_{m(i)}, G_{f(i)}) + \delta^2 \frac{1}{2} V(G_{m(i)}) + Cov(\alpha_{r(m(i))c(m(i))}, \alpha_{r(i)c(i)}) \\ &\quad + Cov(\gamma_{m(i)}^0, \gamma_i^0) + \lambda^2 Cov(SES_{m(i)}, SES_i) \end{aligned}$$

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

$$\begin{aligned} Cov(L_{f(i)}, L_i) &= Cov(\delta G_{f(i)} + \alpha_{r(f(i))c(f(i))} + \gamma_{f(i)}^{s(f(i))} + \lambda SES_{f(i)} + \varepsilon_{f(i)}, \\ &\quad \delta \frac{1}{2}(G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i) \\ Cov(L_{f(i)}, L_i) &= \delta^2 \frac{1}{2} Cov(G_{f(i)}, G_{m(i)}) + \delta^2 \frac{1}{2} V(G_{f(i)}) + Cov(\alpha_{r(f(i))c(f(i))}, \alpha_{r(i)c(i)}) \\ &\quad + \lambda^2 Cov(SES_{f(i)}, SES_i) \end{aligned}$$

The covariance between father and daughter is the same as the covariance between father and son except there is no term for the within-family sex-specific random effect since $Cov(\gamma_i^0, \gamma_j^1) = 0$ for every individual i and j . Similarly, the covariance between mother and son is the same as the covariance between mother and daughter except there is no term for the within-family sex-specific random effect. As such, the same-sex intergenerational coefficient (from regressing a child's lifespan on their parent's lifespan), as reported in Equation (3), is larger than the opposite-sex intergenerational coefficient.

In addition to the covariances between parents and children we can also express the covariances between different types of siblings. Without loss of generality, let i denote one sibling and let j denote the other sibling. Note that the siblings have the same parents such that $m(i) = m(j)$ and $f(i) = f(j)$.

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

$$\begin{aligned} Cov(L_i, L_j) &= Cov(\delta \frac{1}{2}(G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i, \\ &\quad \delta \frac{1}{2}(G_{m(j)} + G_{f(j)} + \eta_j) + \alpha_{r(j)c(j)} + \gamma_j^{s(j)} + \lambda SES_j + \varepsilon_j) \end{aligned}$$

$$Cov(L_i, L_j) = \delta^2 \frac{1}{4} V(G_{m(i)}) + \delta^2 \frac{1}{4} V(G_{f(i)}) + \delta^2 \frac{1}{2} Cov(G_{m(i)}, G_{f(i)}) \\ + Cov(\alpha_{r(i)c(i)}, \alpha_{r(j)c(j)}) + Cov(\gamma_i^1, \gamma_j^1) + \lambda^2 Cov(SES_i, SES_j)$$

Note that if the environment and socioeconomic status of the fathers and sons are the same and the variance of the genetic component is the same for mothers and fathers (i.e., $V(G_{m(i)}) = V(G_{f(i)})$), then the covariance between male siblings is the same as the covariance between father and son. In addition, this implies that if the variance in longevity for fathers and sons is also the same (i.e., $V(L_{f(i)}) = V(L_i)$ for fathers and sons) then the regression coefficient for the intergenerational transmission between fathers and sons will be the same as the regression coefficient between male siblings as reported in Equation (3) of the paper.

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

$$Cov(L_i, L_j) = Cov\left(\delta \frac{1}{2} (G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i, \right. \\ \left. \delta \frac{1}{2} (G_{m(j)} + G_{f(j)} + \eta_j) + \alpha_{r(j)c(j)} + \gamma_j^{s(j)} + \lambda SES_j + \varepsilon_j\right)$$

$$Cov(L_i, L_j) = \delta^2 \frac{1}{4} V(G_{m(i)}) + \delta^2 \frac{1}{4} V(G_{f(i)}) + \delta^2 \frac{1}{2} Cov(G_{m(i)}, G_{f(i)}) \\ + Cov(\alpha_{r(i)c(i)}, \alpha_{r(j)c(j)}) + Cov(\gamma_i^0, \gamma_j^0) + \lambda^2 Cov(SES_i, SES_j)$$

Note that if the environment and socioeconomic status of the mothers and daughters are the same and the variance of the genetic component is the same for mothers and fathers (i.e., $V(G_{m(i)}) = V(G_{f(i)})$), then the covariance between female siblings is the same as the covariance between mothers and daughters. In addition, this implies that if the variance in longevity for mothers and daughters is also the same (i.e., $V(L_{m(i)}) = V(L_i)$ for mothers and daughters) then the regression coefficient for the intergenerational transmission between mothers and daughters will be the same as the regression coefficient between female siblings as reported in Equation (3) of the paper.

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

$$Cov(L_i, L_j) = Cov\left(\delta \frac{1}{2} (G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i, \right. \\ \left. \delta \frac{1}{2} (G_{m(j)} + G_{f(j)} + \eta_j) + \alpha_{r(j)c(j)} + \gamma_j^{s(j)} + \lambda SES_j + \varepsilon_j\right)$$

$$Cov(L_i, L_j) = \delta^2 \frac{1}{4} V(G_{m(i)}) + \delta^2 \frac{1}{4} V(G_{f(i)}) + \delta^2 \frac{1}{2} Cov(G_{m(i)}, G_{f(i)}) \\ + Cov(\alpha_{r(i)c(i)}, \alpha_{r(j)c(j)}) + \lambda^2 Cov(SES_i, SES_j)$$

Note that the difference between male siblings, female siblings, and opposite sex siblings is whether the covariance in the within-family sex-specific random effect is included.

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

$$\begin{aligned} Cov(L_i, L_j) = & Cov\left(\delta\frac{1}{2}(G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i,\right. \\ & \left.\delta\frac{1}{2}(G_{m(j)} + G_{f(j)} + \eta_j) + \alpha_{r(j)c(j)} + \gamma_j^{s(j)} + \lambda SES_j + \varepsilon_j\right) \end{aligned}$$

$$\begin{aligned} Cov(L_i, L_j) = & \delta^2\frac{1}{4}V(G_{m(i)}) + \delta^2\frac{1}{4}V(G_{f(i)}) + \delta^2\frac{1}{2}Cov(G_{m(i)}, G_{f(i)}) \\ & + Cov(\alpha_{r(i)c(i)}, \alpha_{r(j)c(j)}) + Cov(\gamma_i^1, \gamma_j^1) + \lambda^2Cov(SES_i, SES_j) + \delta^2\frac{1}{4}Cov(\eta_i, \eta_j) \end{aligned}$$

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

$$\begin{aligned} Cov(L_i, L_j) = & Cov\left(\delta\frac{1}{2}(G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i,\right. \\ & \left.\delta\frac{1}{2}(G_{m(j)} + G_{f(j)} + \eta_j) + \alpha_{r(j)c(j)} + \gamma_j^{s(j)} + \lambda SES_j + \varepsilon_j\right) \end{aligned}$$

$$\begin{aligned} Cov(L_i, L_j) = & \delta^2\frac{1}{4}V(G_{m(i)}) + \delta^2\frac{1}{4}V(G_{f(i)}) + \delta^2\frac{1}{2}Cov(G_{m(i)}, G_{f(i)}) \\ & + Cov(\alpha_{r(i)c(i)}, \alpha_{r(j)c(j)}) + Cov(\gamma_i^0, \gamma_j^0) + \lambda^2Cov(SES_i, SES_j) + \delta^2\frac{1}{4}Cov(\eta_i, \eta_j) \end{aligned}$$

Note that for twins η_i and η_j are the same. Therefore, the covariances for twins are the same as those for the same sex siblings except that they also have a term for the covariance in the additional shared genetic component. As such, the twin coefficient (from regressing a child's lifespan on their twin's lifespan), as reported in Equation (3), is larger than the same-sex sibling coefficient.

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

$$\begin{aligned} Cov(L_{f(f(i))}, L_i) = & Cov\left(\delta G_{f(f(i))} + \alpha_{r(f(f(i)))c(f(f(i)))} + \gamma_{f(f(i))}^{s(f(f(i)))} + \lambda SES_{f(f(i))} + \varepsilon_{f(f(i))},\right. \\ & \delta\frac{1}{4}(G_{f(f(i))} + G_{m(f(i))} + \eta_{f(i)} + G_{f(m(i))} + G_{m(m(i))} + \eta_{m(i)}) \\ & \left. + \delta\frac{1}{2}\eta_i + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i\right) \end{aligned}$$

$$\begin{aligned} Cov(L_{f(f(i))}, L_i) = & \delta^2\frac{1}{4}V(G_{f(f(i))}) + \delta^2\frac{1}{4}Cov(G_{f(f(i))}, G_{m(f(i))}) + \delta^2\frac{1}{4}Cov(G_{f(f(i))}, G_{f(m(i))}) \\ & + \delta^2\frac{1}{4}Cov(G_{f(f(i))}, G_{m(m(i))}) + Cov(\alpha_{r(f(f(i)))c(f(f(i)))}, \alpha_{r(i)c(i)}) \\ & + Cov(\gamma_{f(f(i))}^1, \gamma_i^1) + \lambda^2Cov(SES_{f(f(i))}, SES_i) \end{aligned}$$

Note, that $f(f(i))$ is individual i 's father's father (paternal grandfather), $f(m(i))$ is individual

i 's mother's father (maternal grandfather), $m(m(i))$ is individual i 's mother's mother (maternal grandmother), and $m(f(i))$ is individual i 's father's mother (paternal grandmother). Each of the covariances for the other three grandparent-grandchild combinations are analogous to this results, however, with the covariance of the within-family sex-specific random effect omitted for the cross-sex covariances.

The covariance between the average lifespan of parents and son is given by:

$$\begin{aligned} Cov(\frac{1}{2}L_{f(i)} + \frac{1}{2}L_{m(i)}, L_i) &= Cov(\frac{1}{2}[\delta G_{f(i)} + \alpha_{r(f(i))c(f(i))} + \gamma_{f(i)}^{s(f(i))} + \lambda SES_{f(i)} + \varepsilon_{f(i)}] + \frac{1}{2}[\delta G_{m(i)} + \alpha_{r(m(i))c(m(i))} \\ &\quad + \gamma_{m(i)}^{s(m(i))} + \lambda SES_{m(i)} + \varepsilon_{m(i)}], \delta \frac{1}{2}(G_{m(i)} + G_{f(i)} + \eta_i) + \alpha_{r(i)c(i)} + \gamma_i^{s(i)} + \lambda SES_i + \varepsilon_i) \end{aligned}$$

$$\begin{aligned} Cov(\frac{1}{2}L_{f(i)} + \frac{1}{2}L_{m(i)}, L_i) &= \delta^2 \frac{1}{2} Cov(G_{f(i)}, G_{m(i)}) + \delta^2 \frac{1}{4} V(G_{f(i)}) + \delta^2 \frac{1}{4} V(G_{m(i)}) \\ &\quad + \frac{1}{2} Cov(\alpha_{r(f(i))c(f(i))}, \alpha_{r(i)c(i)}) + \frac{1}{2} Cov(\alpha_{r(m(i))c(m(i))}, \alpha_{r(i)c(i)}) + \frac{1}{2} Cov(\gamma_{f(i)}^1, \gamma_i^1) \\ &\quad + \frac{1}{2} \lambda^2 Cov(SES_{f(i)}, SES_i) + \frac{1}{2} \lambda^2 Cov(SES_{m(i)}, SES_i) \end{aligned}$$

Note that to obtain the regression coefficient when regressing son lifespan on the average lifespan of parents, this covariance will be divided by the variance of average parent lifespan, $V(\frac{1}{2}L_{f(i)} + \frac{1}{2}L_{m(i)})$ instead of $V(L_{f(i)})$. Since $V(\frac{1}{2}L_{f(i)} + \frac{1}{2}L_{m(i)}) = \frac{1}{4}V(L_{f(i)}) + \frac{1}{4}V(L_{m(i)}) + \frac{1}{2}Cov(L_{f(i)}, L_{m(i)})$ and the correlation between father and mother adult lifespan is quite low (see Table A.4), this denominator is likely much smaller than $V(L_{f(i)})$ (the denominator when regressing son lifespan on father lifespan).

2 Regression Coefficient of Son's Lifespan on Father's Lifespan

The regression coefficient of son's lifespan on father's lifespan, denoted by β_L , is given by:

$$\beta_L = \frac{\delta^2 \frac{1}{2} Cov(G_{f(i)}, G_{m(i)}) + \delta^2 \frac{1}{2} Var(G_{f(i)}) + Cov(\alpha_{r(f(i))c(f(i))}, \alpha_{r(i)c(i)}) + Cov(\gamma_{f(i)}^1, \gamma_i^1) + \lambda^2 Cov(SES_{f(i)}, SES_i)}{Var(L_{f(i)})}$$

Suppose now that the intergenerational transmission of SES is given by:

$$SES_i = \beta_{SES} SES_{f(i)} + \epsilon_i$$

where β_{SES} is the intergenerational SES transmission coefficient. Then the regression coefficient of son's lifespan on father's lifespan, denoted by β_L , is given by:

$$\beta_L = \frac{\delta^2 \frac{1}{2} \text{Cov}(G_{f(i)}, G_{m(i)}) + \delta^2 \frac{1}{2} \text{Var}(G_{f(i)}) + \text{Cov}(\alpha_{r(f(i))c(f(i))}, \alpha_{r(i)c(i)}) + \text{Cov}(\gamma_{f(i)}^1, \gamma_i^1) + \lambda^2 \beta_{SES} \text{Var}(SES_{f(i)})}{\text{Var}(L_{f(i)})}$$

Therefore, as written in Equation (4) of the paper:

$$\beta_L = \lambda^2 \frac{\text{Var}(SES_{f(i)})}{\text{Var}(L_{f(i)})} \beta_{SES} + \pi$$

where

$$\pi = \frac{\delta^2 \frac{1}{2} \text{Cov}(G_{f(i)}, G_{m(i)}) + \delta^2 \frac{1}{2} \text{Var}(G_{f(i)}) + \text{Cov}(\alpha_{r(f(i))c(f(i))}, \alpha_{r(i)c(i)}) + \text{Cov}(\gamma_{f(i)}^1, \gamma_i^1)}{\text{Var}(L_{f(i)})}$$

3 Covariance of Indirect Utility V_c and V_{c-1}

To keep expressions as simple as possible, we consider the case in which the indirect utility of generation c is given by:

$$V_c = V_c(Y, S) = \text{rank}_c \times LE_c$$

The covariance between the indirect utilities of generations c and $c - 1$ is:

$$\text{Cov}(V_c, V_{c-1}) = \text{Cov}(\text{rank}_c LE_c, \text{rank}_{c-1} LE_{c-1})$$

Applying the product rule of covariance, we obtain:

$$\begin{aligned} &= \mathbb{E}(\text{rank}_c) \mathbb{E}(\text{rank}_{c-1}) \text{Cov}(LE_c, LE_{c-1}) + \mathbb{E}(\text{rank}_c) \mathbb{E}(LE_{c-1}) \text{Cov}(\text{rank}_c, \text{rank}_{c-1}) \\ &+ \mathbb{E}(LE_c) \mathbb{E}(\text{rank}_{c-1}) \text{Cov}(\text{rank}_c, LE_{c-1}) + \mathbb{E}(LE_c) \mathbb{E}(LE_{c-1}) \text{Cov}(\text{rank}_c, \text{rank}_{c-1}) \\ &+ \text{Cov}(\text{rank}_c, \text{rank}_{c-1}) \text{Cov}(LE_c, LE_{c-1}) + \text{Cov}(\text{rank}_c, LE_{c-1}) \text{Cov}(LE_c, \text{rank}_{c-1}) \end{aligned}$$

This expression shows that even in this simplest case, the lifetime covariances are not a simple

linear combination of the lifespan and rank covariances. They depend also on, for example, the covariance between the parent's SES and the child's lifespan.