Sibling Similarity in Old Age Mortality

Federica Querin* Benedetta Scotti** Nicola Barban*

Abstract

It is well established that childhood family conditions play an important role in shaping survival chances around birth and in adult life. Still, it is an open question whether the influence of family of origin on old age mortality patterns holds in population-wide settings, especially in low mortality contemporary countries. To the best of our knowledge, ours is the first study to explore similarity in sibling mortality among seniors at the population level in the United States, a low mortality but high inequality country. We combine data from multiple US administrative data sources, including the 1920 and 1940 census, and CenSoc DMF dataset, which includes death records from the Social Security Administration Death Master File. We compare sibling dyads and matched dyads who share parental and background characteristics following the methodology suggested in Raab et al. (2014) for life course trajectories. Findings suggest that siblings' lifespans are more similar than that of matched-but-unrelated dyads, with persisting lifespan correlations at older ages. We find no evidence of socio-economic differentials in sibling correlations by father's occupation, rural or urban setting, and county of residency in childhood.

Theoretical Framework

Siblings tend to have similar life course trajectories when it comes to a number of socio-economic outcomes (Karhula et al., 2019; Raab et al., 2014). There is also evidence of siblings' correlation in both child (Curtis et al., 1993; Das Gupta, 1990; Guo, 1993; Scalone et al., 2017) and adult mortality (Alter et al., 2001; Kröger et al., 2018; Minardi et al., 2024), ascribable to shared genetic and environmental factors. There is also a large demographic literature on the impact of childhood conditions on adult mortality, the so-called "long arm of childhood" (Baranowska-Rataj et al., 2017; Hayward and Gorman, 2004; Kröger et al., 2018; van Dijk et al., 2019). However, fewer studies take a sibling-approach to establish lasting family influences at older ages (Alter et al., 2001; Kröger et al., 2018; Minardi et al., 2024), and most of them are based on historical populations or low- and middle-income countries with high mortality as discussed in Baranowska-Rataj et al. (2017).

This paper investigates whether there are family influences on later life mortality in the United States, a low-mortality country characterized by high levels of inequality, and whether they manifest differently along are socio-economic lines. Focusing on individuals who are at or past retirement age (i.e., at least 65 years old) allows us to shed some light on a crucial demographic group who is living at ages around current life expectancy in low mortality countries.

* University of Bologna; ** Cassa Depositi e Prestiti. Corresponding author: federica.querin@unibo.it

Given sibling correlations in life course trajectories, child mortality, and adult mortality ascribable to shared genetic and environmental conditions experienced early in life, we posit that mortality correlation among siblings persists into old age too (H1a). However, the influence of shared family background may wane as individuals age, and different behavioural and contextual factors may take over. Thus, the alternative hypothesis is that siblings do not share more similar old-age mortality profiles compared to dyads of unrelated individuals (H1b).

Sibling correlations may be heterogeneous along several dimensions, including socio-economic background, geography, and birth cohorts. Previous research has shown that individuals in upper socio-economic strata tend to die at more similar ages compared to individuals with lower socio-economic status (Sasson, 2016). We may thus expect siblings from upper class families to die at more similar ages compared to individuals from lower class families (HP2a). However, low mortality countries display a lower variation in age at death, which may lead to a reduction in lifespan heterogeneity across groups. Therefore, the competing hypothesis is that at the population-level there are no identifiable differences in the family effects across socio-economic groups (HP2b).

To adjudicate these hypotheses and answer our research questions we use population-wide data for the United States through census data and administrative death records. We construct three analytical samples: (i) sibling dyads, who naturally share genetic, parental background characteristics, and early life environmental conditions; (2) matched dyads, who are constructed to have same parental socio-economic characteristics and some common environmental conditions; (3) random dyads, who are two randomly selected individuals from the sample population and constitute a benchmark reference.

Data and Methods

We combine data from multiple US administrative data sources. First, data from both the complete 1920 and the 1940 US Federal Census are accessible through the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al., 2020). The earlier census allows us to have a population-wide representation of individuals who were born between 1905 and 1919 and were therefore registered as children in the 1920 US Federal Census. Following Kröger et al. (2018), we consider all children registered in the same household in the 1920 census as siblings. Second, we use data from the CenSoc project (Goldstein et al., 2021) for information on age at death. This project links all deaths reported to the US Social Security Administration between 1975 and 2005 with the 1940 census data, providing a reliable large-scale microdata set for contemporary mortality in the United States. More specifically, the CenSoc-DMF links a collection of over 83 million death records from the Death Master File (DMF) to the 1940 census relying on first name, last name, and year of birth (Abramitzky et al., 2020). While this dataset is the most comprehensive publicly available for the United States, it has the major limitation of only including males. Therefore, our sibling pairs consists solely of brother-brother dyads, a limitation which we further discuss where appropriate.

Table A.1 synthetizes the sample construction process. Our final sample consists of 296,209 men, clustered into 139,785 households of origin. For each individual we know family characteristics (from 1920 census), individual-level characteristics (from 1940 census), and date of death (from the social security administration death master file via CenSoc).

Dyads Construction

Adapting the analytical strategy developed by Raab et al. (2014) to study sibling similarity in life course outcomes, we construct sibling and matched dyads, and add random dyads as an additional control group. We generate sibling dyads by matching each individual in the sample with his sibling(s). For individuals with only one brother in the sample, there is only possible match. For individuals with two or more brothers in the sample, multiple matches are possible, but we keep only one random dyad per family. This results in 139,333 sibling dyads. To generate unrelated dyads, we first match each individual in the sample to all individuals sharing the same family background characteristics, measured by year of birth, county (and state) of residence, urban/rural residency, and father's occupation in the 1920 Census. We exclude dyads so generated where individuals are, in fact, siblings. Out of all possible matches, we then keep one randomly chosen match per individual. This procedure yields 155,545 matched dyads. Lastly, we generate 293,365 random dyads consisting in two individuals in the sample who were born in the same county but are otherwise not matched on any other observable characteristic.

Analytical framework

To assess the correlation between age at death among siblings we regress the age at death of one member of a dyad on the age at death of the other, distinguishing between the three types of relationships (siblings, matched, random) as follows:

$$\log(Y_i) = D_{ij} + D_{ij} * log(Y_j) + BY_i + County_i + BY_j + County_j + X_i^{1920} + X_j^{1920} + X_i^{1940} + X_j^{1940} + \varepsilon_{ij}$$

$Y_{i,j}$ is the age at death (logged) of the two individuals in each dyad (indicated with i and j, respectively). The main variable of interest is D_{ij}, which represents the three possible relationship-status within the dyads: (1) siblings, if both were children belonging to the same household in the 1920 census; (2) matched dyads, for individuals who have the same age gap as an in-sample sibling dyad, were born in the same county, had the same rural or urban residency status, and had a father with the same HISCO classification; (3) random dyads, for two individuals in the final sample who were born in the same county but do not necessarily share other characteristics. Additional controls include the exact year of birth for both individuals in the dyad, their county of residence, family characteristics measured in 1920, and individual-level characteristics measured in 1940. Family characteristics include all the matching variables plus birth year of mother and father, and the number of siblings presents in the household in 1920. Individual characteristics collected in 1940 include state of residence, marital status, race, education level, broad occupational category, and urban or rural residency status. These may vary also between siblings once they leave the parental home.

Results

Table 1 presents descriptive information for all variables included in the models separately by dyad relationship status (siblings, matched, random) and for each member of the dyad. Overall, the characteristics are similar across the three groups and within members of the dyads. On average, individuals in the sample were born around 1912 and their mothers were born around 1883, making them around 29 years of age at the time of births. Fathers were on average four years older than the mothers, and there are on average 5 children in the household in 1920 across all groups. Urbanicity is one of the significantly different characteristics by relationship type. Indeed, while for sibling and random dyads about 40.7% of the sample lives in an urban setting, almost 53% of the matched dyad do so. The other difference is in the proportion of dyads with a father's top occupational characteristic, which is lower for matched dyad given the difficulty of finding an exact match in a relatively smaller pool. For all sub-groups, the most common father's occupation is self-employed (farmers and fishermen), followed by skilled workers (medium and lower), and unskilled workers, with a fifth of the information missing. The sample is almost exclusively white.

The main outcome of interest is the (log) difference at age at death comparing sibling dyads with the other two control dyads. Figure 1 presents results for the difference in age at death in years. Overall, given that individuals belong to the same birth cohorts (1905-1919), the deaths are concentrated within three years of each other, although there is a long right tail. Sibling dyads, indicated in blue, tend to die within ten years of each other more than unrelated dyads. After that, there is a crossover, suggesting that at the population level siblings are more similar in lifespan when compared with a control group of matched and random dyads. Interestingly, there are no substantial differences between the matched and the random dyads, despite matched dyads sharing important characteristics.

Figure 2 reports the marginal effects for the main model from eq. (1) with additive controls. The base model (M0, in blue) simply estimates the association between the log ages at death of the two individuals in a dyad, separately by relationship type. Both sibling and matched dyads show positive and significant correlations, albeit the size is quite smaller for matched dyads. Random individuals' ages at death are not associated even in this simple model. The subsequent model (M1, in red) adds year of birth and state fixed effects, which reduces the size of the correlation, as expected. When adding the family characteristics listed in Table A.2 (M2, in green), the correlation further decreases, and it is no longer statistically distinguishable from zero for matched dyads. The last model (M3, in yellow) also includes individual characteristics that are found to be important mediators in previous literature. This marginally reduces the correlations for siblings, which remains positive and significant. The permanence of a positive correlation for siblings, but crucially not for matched and random dyads, supports H1a in that mortality correlation among siblings persists in old age and cannot be explained away by including observable family and individual characteristics.

Our second research question explores whether the sibling correlations in later life mortality change across parental socio-economic status. To answer it, we focus on the difference between siblings and matched dyads and run a fully interacted model with dyad type (sibling or matched) and fathers' occupation. Figure 3 reports the results for sibling dyads, which are *de facto* the correlation in lifespan found in Figure 2 for siblings in the full model (M3) stratified by father's occupation. Indeed, all the estimations are around the 4-7% range. While there may appear to be a slight negative gradient, all confidence intervals partially overlap, meaning that we do not find evidence of different correlations in older age mortality within siblings depending on parental socio-economic conditions in childhood (HP2b supported). Moreover, we confirm that the absence of correlations for matched dyads is not driven by a specific father's occupation, but they are all null (Figure A.1). Figure A.2 reports results for siblings and matched dyads by urban status. Similarly to father's occupation, we do not observe heterogeneities by rural or urban setting. Likewise, there is not a clear geographical pattern that emerges from a spatial analysis.

Figures and Tables

Figure 1 – Difference in Age at Death in Years between Siblings, Matched, and Random Dyads

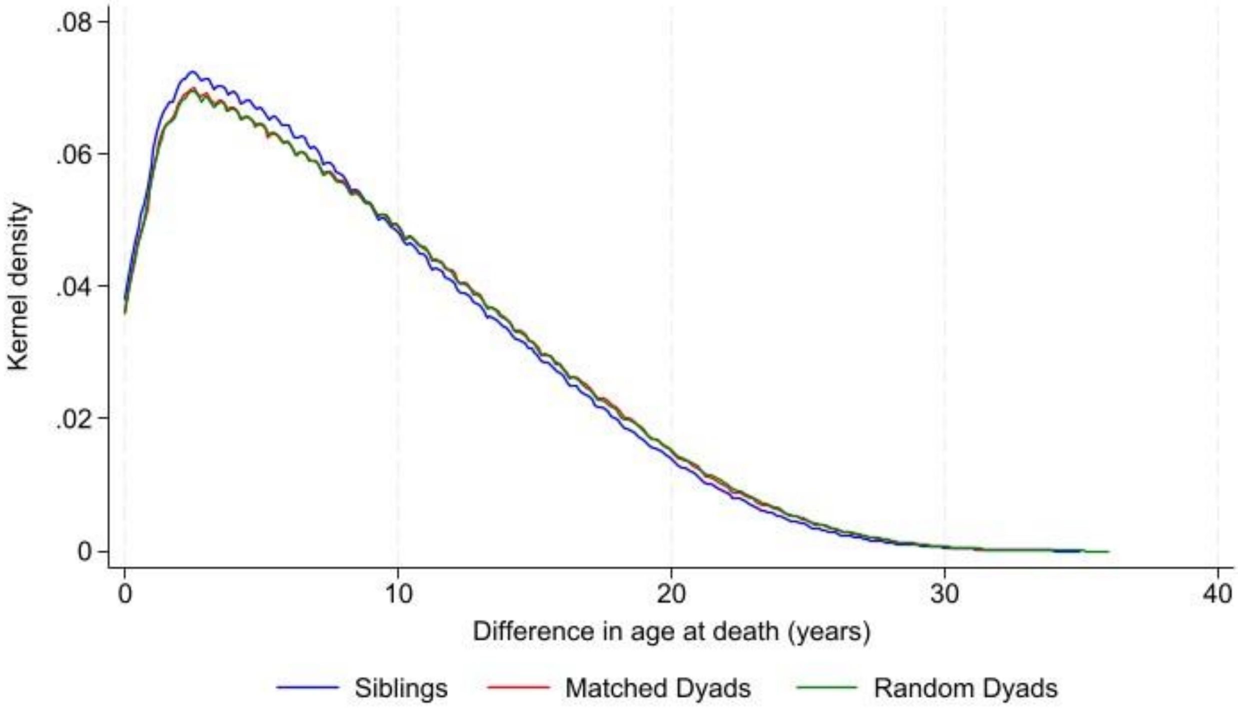


Figure 2 – Correlations in Lifespan by Dyad Relationship Type

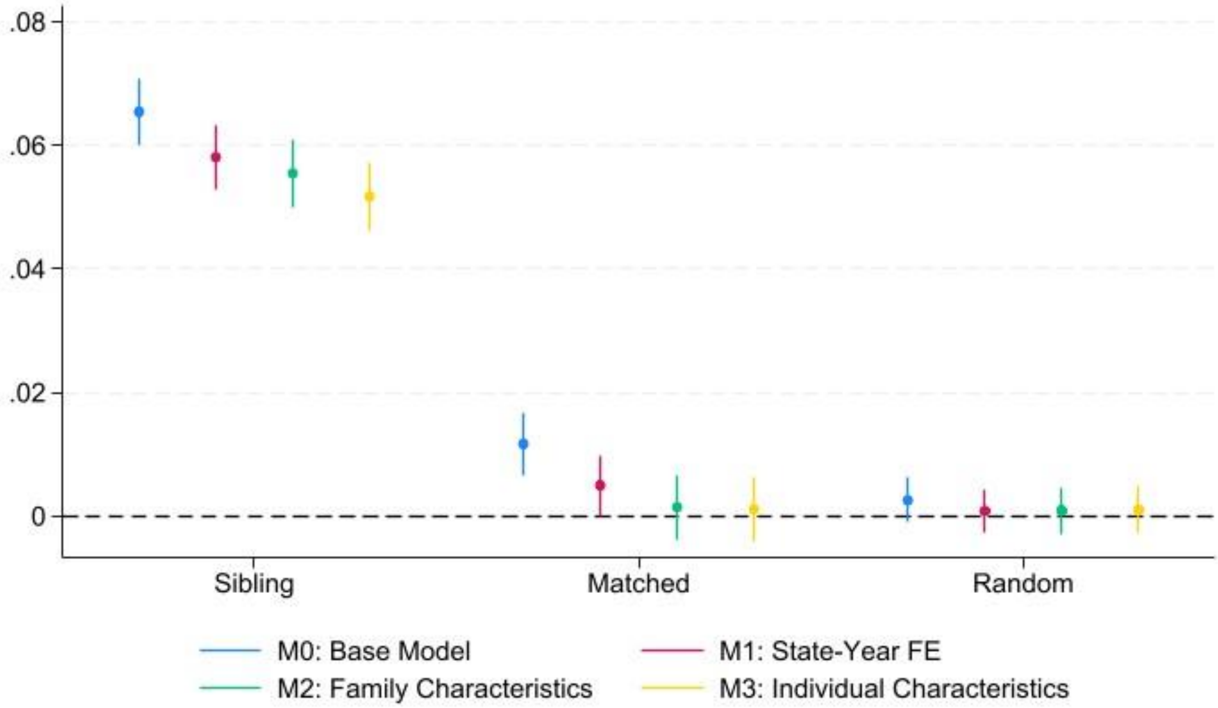
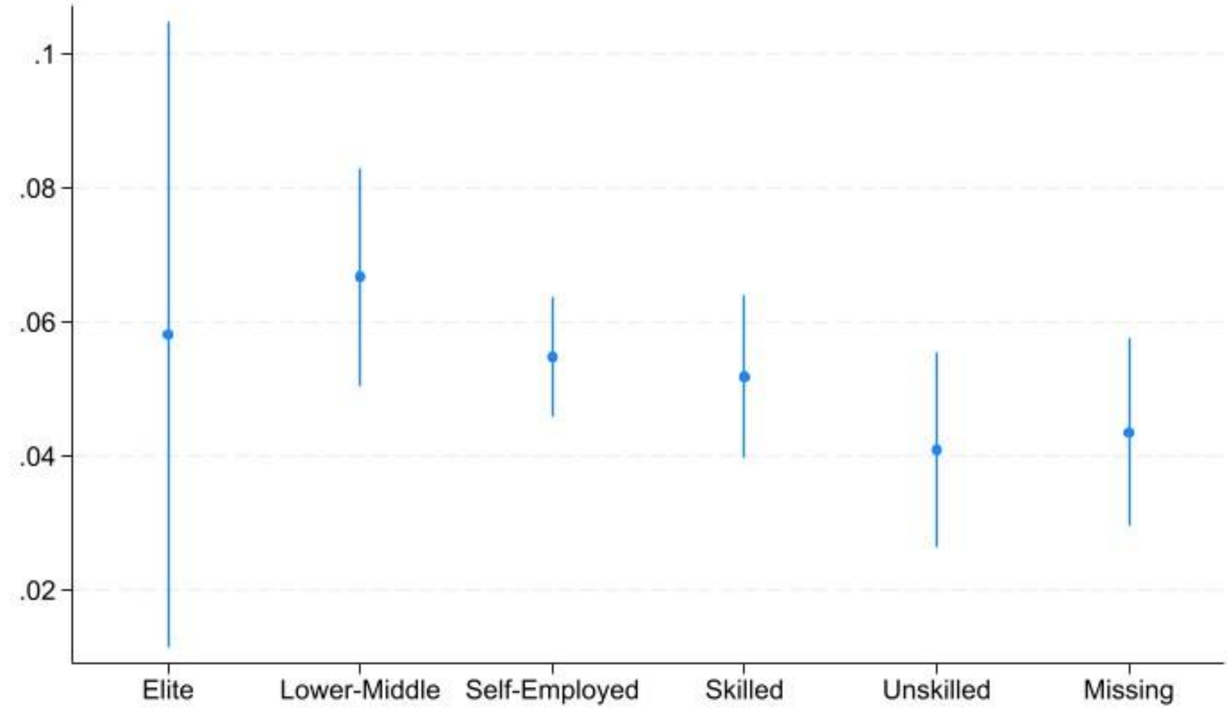


Figure 3 – Correlations in Lifespan by Father's Occupation, Siblings Only



Notes: Margins from three-way interaction with sibling dyad (sibling or matched), continuous (log)age at death, and father's occupation. For ease of presentation, results for matched dyads are reported in Figure A.1. Father's occupation is classified as elite, lower middle class, self-employed (farmers and fishermen), skilled workers (medium and lower), unskilled workers, missing occupation.

Table 1 – Descriptive Statistics by Dyad Relationship Status

	Siblings	Matched	Random
Birth Year i	1912.163 (3.937)	1912.190 (3.879)	1912.158 (3.936)
Birth Year j	1912.163 (3.931)	1912.222 (3.815)	1912.156 (3.934)
Death Age i	79.011 (7.552)	79.018 (7.573)	79.035 (7.558)
Death Age j	79.018 (7.565)	79.023 (7.575)	79.003 (7.562)
Mother's birth year i	1883.571 (7.042)	1883.536 (6.972)	1883.530 (6.999)
Mother's birth year j	=	1883.565 (6.944)	1883.574 (7.034)
Father's birth year i	1878.974 (8.159)	1879.102 (7.936)	1878.913 (8.118)
Father's birth year j	=	1879.148 (7.869)	1878.952 (8.116)
1920			
Urban i	40.66%	52.99%	40.66%
Urban j	=	=	40.82%
Number siblings i	5.164 (2.108)	5.168 (2.130)	5.229 (2.216)
Number siblings j	=	5.134 (2.123)	5.156 (2.117)
Father's Occupation i			
Elite	1.27%	0.37%	1.25%
Lower Middle	10.67%	9.8%	10.61%
Self-Employed	36.19%	36.24%	36.3%
Skilled	19%	21.88%	19.01%
Unskilled	12.95%	11.75%	13%
Missing	19.92%	19.96%	19.84%
Father's Occupation j	=	=	
Elite			1.25%
Lower Middle			10.66%
Self-Employed			36.07%
Skilled			19.1%
Unskilled			13.03%
Missing			19.89%
1940			
Race i: white	98.19%	98.97%	98.33%
Race j: white	98.20%	98.95%	98.30%
Urban i	51.66%	58.89%	51.64%
Urban j	51.76%	58.78%	51.79%
Education i			
<4th Grade	2.71%	1.91%	2.64%
Grades 5-8	38.15%	39.18%	38.61%
Grades 9-11	22.47%	23.21%	22.32%
Grade 12	23.21%	22.61%	23.09%

High School +	11.92%	11.56%	11.77%
Missing	1.53%	1.53%	1.56%
Education j			
<4th Grade	2.76%	1.87%	2.62%
Grades 5-8	38.28%	38.99%	38.35%
Grades 9-11	22.29%	23.22%	22.37%
Grade 12	23.07%	22.74%	23.19%
High School +	12.03%	11.64%	11.94%
Missing	1.57%	1.54%	1.53%
Own Occupation i			
Professional	5.47%	5.56%	5.43%
Farmers	11.59%	10.67%	11.64%
Managers, Officials	5.66%	5.60%	5.67%
Clerical	8.11%	9.42%	8.12%
Sales	4.37%	4.67%	4.31%
Craftsmen	12.49%	12.97%	12.50%
Operatives	21.84%	22.64%	21.85%
Service	3.88%	3.80%	3.83%
Laborers	20.83%	18.85%	20.87%
Missing	5.75%	5.81%	5.78%
Own Occupation j			
Professional	5.53%	5.65%	5.51%
Farmers	11.65%	10.71%	11.57%
Managers, Officials	5.75%	5.54%	5.68%
Clerical	8.26%	9.34%	8.19%
Sales	4.31%	4.64%	4.34%
Craftsmen	12.42%	12.89%	12.48%
Operatives	21.67%	22.82%	21.74%
Service	3.81%	3.84%	3.84%
Laborers	20.79%	18.90%	20.89%
Missing	5.82%	5.66%	5.76%
Marital Status i			
Never Married	38.09%	40.45%	38.19%
Married	58.97%	56.86%	58.87%
Other	2.93%	2.69%	2.94%
Marital Status j			
Never Married	37.92%	40.25%	38.08%
Married	59.08%	57.06%	58.97%
Other	3%	2.69%	2.95%

Notes: Authors' calculation from 1920 and 1940 US Census. Categorical variables indicated as percentage of the sample in the indicated categories, which sum to 1 save rounding. Means for all other variables, standard deviation in parentheses. Equal sign (=) indicates that for individual j that characteristic is the same as for individual i.

References

Abramitzky, R., Mill, R., & Pérez, S. (2020). Linking individuals across historical sources: A fully automated approach. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *53*(2), 94–111. https://doi.org/10.1080/01615440.2018.1543034

Alter, G., Oris, M., & Broström, G. (2001). The Family and Mortality: A Case Study from Rural Belgium: *Annales de Démographie Historique*, *101*(1), 11–31. https://doi.org/10.3917/adh.101.0010

Baranowska-Rataj, A., Barclay, K., & Kolk, M. (2017). The effect of number of siblings on adult mortality: Evidence from Swedish registers for cohorts born between 1938 and 1972. *Population Studies*, *71*(1), 43–63. https://doi.org/10.1080/00324728.2016.1260755

Black, S., Duzett, N., Lleras-Muney, A., Pope, N., & Price, J. (2023). *Intergenerational Transmission of Lifespan in the US* (w31034; p. w31034). National Bureau of Economic Research. https://doi.org/10.3386/w31034

Curtis, S. L., Diamond, I., & McDonald, J. W. (1993). Birth interval and family effects on postneonatal mortality in Brazil. *Demography*, *30*(1), 33–43. https://doi.org/10.2307/2061861

Das Gupta, M. (1990). Death Clustering, Mothers' Education and the Determinants of Child Mortality in Rural Punjab, India. *Population Studies*, *44*(3), 489–505. https://doi.org/10.1080/0032472031000144866

Goldstein, J. R., Alexander, M., Breen, C., González, A. M., Menares, F., Osborne, M., Snyder, M., & Yildirim, U. (2021). *CenSoc Mortality File: Version 2.0* (Version 2.0) [Dataset]. Berkeley: University of California.

Guo, G. (1993). Use of Sibling Data to Estimate Family Mortality Effects in Guatemala. *Demography*, *30*(1), 15. https://doi.org/10.2307/2061860

Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography*, *41*(1), 87–107. https://doi.org/10.1353/dem.2004.0005

Karhula, A., Erola, J., Raab, M., & Fasang, A. (2019). Destination as a process: Sibling similarity in early socioeconomic trajectories. *Advances in Life Course Research*, *40*, 85–98. https://doi.org/10.1016/j.alcr.2019.04.015

Kröger, H., Hoffmann, R., Tarkiainen, L., & Martikainen, P. (2018). Comparing Observed and Unobserved Components of Childhood: Evidence From Finnish Register Data on Midlife Mortality From Siblings and Their Parents. *Demography*, *55*(1), 295–318. https://doi.org/10.1007/s13524-017-0635-6

Minardi, S., Corti, G., & Barban, N. (2024). Historical Patterns in the Intergenerational Transmission of Lifespan and Longevity: A Research Note on U.S. Cohorts Born Between 1700 and 1900. *Demography*, *61*(4), 979–994. https://doi.org/10.1215/00703370-11458359

Perls, T. T., Wilmoth, J., Levenson, R., Drinkwater, M., Cohen, M., Bogan, H., Joyce, E., Brewster, S., Kunkel, L., & Puca, A. (2002). Life-long sustained mortality advantage of siblings of centenarians. *Proceedings of the National Academy of Sciences*, *99*(12), 8442–8447. https://doi.org/10.1073/pnas.122587599

Piraino, P., Muller, S., Cilliers, J., & Fourie, J. (2014). The transmission of longevity across generations: The case of the settler Cape Colony. *Research in Social Stratification and Mobility*, *35*, 105–119. https://doi.org/10.1016/j.rssm.2013.08.005

Raab, M., Fasang, A. E., Karhula, A., & Erola, J. (2014). Sibling Similarity in Family Formation. *Demography*, *51*(6), 2127–2154. https://doi.org/10.1007/s13524-014-0341-6

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2020). *IPUMS USA* (Version 10.0) [Dataset]. Minneapolis, MN: Minnesota Population Center.

Sasson, I. (2016). Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010. *Demography*, *53*(2), 269–293. https://doi.org/10.1007/s13524-015-0453-7

Scalone, F., Agati, P., Angeli, A., & Donno, A. (2017). Exploring unobserved heterogeneity in perinatal and neonatal mortality risks: The case of an Italian sharecropping community, 1900–39. *Population Studies*, *71*(1), 23–41. https://doi.org/10.1080/00324728.2016.1254812

Smith, K. R., Mineau, G. P., Garibotti, G., & Kerber, R. (2009). Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Social Science & Medicine*, *68*(9), 1649–1658. https://doi.org/10.1016/j.socscimed.2009.02.010

Van Dijk, I. K. (2019). Early-life mortality clustering in families: A literature review. *Population Studies*, *73*(1), 79–99. https://doi.org/10.1080/00324728.2018.1448434

Van Dijk, I. K., Janssens, A., & Smith, K. R. (2019). The Long Harm of Childhood: Childhood Exposure to Mortality and Subsequent Risk of Adult Mortality in Utah and The Netherlands. *European Journal of Population*, *35*(5), 851–871. https://doi.org/10.1007/s10680-018-9505-1

Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations*. *Demography*, *36*(4), 475–495. https://doi.org/10.2307/2648085

Appendix

Table A.1 – Sample Construction

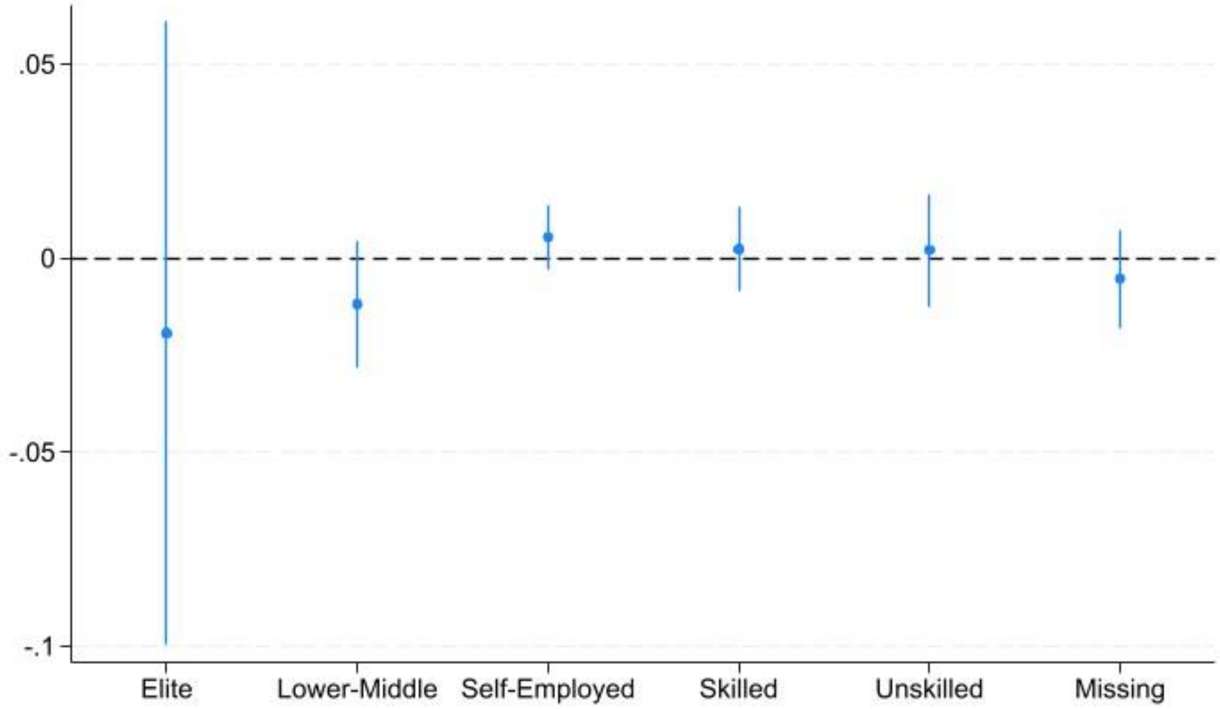
	Individuals	Household of origin
Initial Sample (male children born 1905-1909 in the US 1920 Census)	15,530,409	9,293,163
After linkage with 1940 census	5,691,445	4,421,633
After linkage with the CenSoc DMF dataset	1,574,153	1,400,124
After keeping individuals dying at 65+	1,478,677	1,322,253
After keeping individuals with at least one brother left in the sample	296,209	139,578

Table A.2 – Control Variables

	Used for Matching	Family Characteristics	Individual Characteristics	Varies in sibling dyad	Varies in matched dyad	Varies in random dyad
1920						
Birth year	x		x			x
County	x	x				
Rural/Urban	x	x				x
Father's Occupation	x	x				x
Mother's birth year		x			x	x
Father's birth year		x			x	x
Number of siblings		x			x	x
1940						
State			x	x	x	x
Marital status			x	x	x	x
Education			x	x	x	x
Race		x			x	x
Rural/Urban			x	x	x	x
Occupation			x	x	x	x

Notes: Synopsis of control variables. State of residence is implicit in county of residence. Father's occupation is classified as elite, lower middle class, self-employed (farmers and fishermen), skilled workers (medium and lower), unskilled workers, missing occupation. Marital status is classified as married, never married, other (e.g., widower or separated). Education's categories: less than 4th grade, 5-8 grade, 9-11 grade, 12th grade (corresponds to high school graduation), more than high school, missing. Race: dummy whether white or not. Own occupation in ten categories: professional, famers and farm mangers; proprietors, managers, officials (no farm); clerical; sales; craftsmen; operatives; service; labourers (including farm); missing.

Figure A.1 – Correlations in Lifespan by Father's Occupation, Matched Dyads Only



Notes: Margins from three-way interaction with sibling dyad (sibling or matched), continuous (log)age at death, and father's occupation. Father's occupation is classified as elite, lower middle class, self-employed (farmers and fishermen), skilled workers (medium and lower), unskilled workers, missing occupation.

Figure A.2 – Correlations in Lifespan by Urban or Rural Residence, Siblings and Matched Dyads