Sibling Similarity in Old Age Mortality

Federica Querin* Benedetta Scotti** Nicola Barban*

Introduction

It is well established that family of origin plays an important role in shaping survival chances around birth (van Dijk 2019) and in adult life (Kröger et al. 2017). Less is known about whether, and to what extent, family of origin shapes survival chances in old age. A few studies, based on restricted pedigree data, find that siblings of centenarians have a significant mortality advantage compared to the average population which persists into old age (Perls et al. 2002). Still, whether the influence of family of origin in old age mortality patterns holds in population-wide settings is an open question. To the best of our knowledge, ours is the first study to explore similarity in sibling mortality among seniors in a population-wide settings. In doing so, this work contributes to the understanding of both mortality inequalities and intergenerational transmission of longevity. We compare sibling dyads and unrelated dyads who share parental and background characteristics following the methodology suggested in Raab et al. (2014). To do so, we combine data from multiple US administrative data sources, including the 1920 and 1940 census (Ruggles et al. 2020), CenSoc DMF dataset (Goldstein et al. 2021), and the Social Security Administration Death Master file (SSDM). Preliminary results suggest that siblings' longevity is more similar than that of matched-but-unrelated dyads.

Theoretical framework

Our main research question reads as follows: do siblings share mortality patterns in old age? On the one hand, we know 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. 2017), ascribable to shared genetic and environmental factors. We may thus posit 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 take over. Thus, the alternative hypothesis is that siblings do not share more similar old-age mortality profiles compared to pairs of unrelated persons (H1b).

* University of Bologna; ** Cassa Depositi e Prestiti

We expect heterogenous effects on several dimensions, including socio-economic background, geography, and birth cohorts. For example, 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 dies at more similar ages compare do individuals from lower class families (HP2).

Data

We combine data from multiple US administrative data sources. Sample construction involved the following steps. First, we selected all male individuals born between 1905 and 1919 registered as children in the 1920 US Federal Census (15,530,409 individuals, belonging to 9,293,163 households). We consider children registered in the same household in the 1920 census as siblings. We then link individuals belonging to this subset of the 1920 census to the 1940 US Federal Census. 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). From the 1920 census, we keep information about household county and state of residence, parents' occupation and cohort of birth, and urban/rural status. From the 1940 census, we keep information about focal individuals' county and state of residence, education level, and occupation. Census linkage was performed using the 1920-1940 census crosswalk available through the Census Linking Project (Abramitzky et al. 2020). About 37% of individuals from the 1920 census sample were successfully linked to the 1940 Census[1], leaving us with a sample of 5,691,445 individuals, nested into 4,421,633 households of origin (i.e., households which individuals belonged to as reported by the 1920 census). Out of these, about 28% were further linked to the CenSoc DMF dataset (Goldstein et al. 2021). The latter links the 1940 census to the Social Security Administration Death Master file (SSDM), which keeps track of all deaths that have been reported to the Social Security Administration between 1975 and 2005. After linkage to the CenSoc-DMF dataset, we are left with a sample of 1,574,153 male individuals. Out of these, we keep individuals (i) who die aged 65+, as 65 is the age by which individuals in the observed period could be reasonably expected to start collecting pension benefits, and (ii) who have at least one sibling left in the sample after all these steps. Our final sample consists of 296,209

---

[1] We dropped individuals with multiple linkages resulting from different linkage criteria.

individuals, clustered into 139,785 households of origin, defined based on the 1920 census. Table 1 synthetizes the sample construction process.

4 Methods

4.1 Analytical framework

Our primary objective is to study sibling similarity in old age mortality. Building on previous works, a standard approach would be that of estimating siblings' correlation in mortality by means of survival models with heterogeneous effects at the family of origin level (Alter et al. 2001; Kröger et al. 2018). However, the selected nature of our sample makes such an approach problematic. The most relevant challenge stems from the doubly truncated nature of CenSoc-DMF data, which does not allow to use standard survival analysis techniques. While solutions have been developed to account for double truncation in survival analysis (Mandel et al. 2018; Rennert and Xie 2018), they generally assume independence of survival and truncation times. In our case, such assumption is unlikely to hold since truncation times depend on factors, such as birth cohort, which are determinants of survival times too.

To partially circumvent the selection issue posed by doubly truncated mortality data, we adopt an approach whereby we compare mortality profiles of sibling pairs in our final sample with those of randomly matched dyads of unrelated individuals (pseudo-siblings) extracted from the very same sample. Drawing from the approach developed by Raab et al. (2014) to study sibling similarity in life course outcomes, we use a dyadic linear regression framework with dyadic difference in age-at-death measured in months (logged) as dependent variable and sibling status of the dyad as focal independent variable.[2] As further detailed below, we match unrelated persons so as to ensure similarity of observable family background characteristics. Matching is performed using the following variables: mother birth cohort, county of residence, urban/rural status, and father occupation as measured in the 1920 census. Conditional random matching allows to equalize variance in observed family background characteristics for sibling and unrelated dyads. The main specification looks as follows:

$$\log(|\Delta_d|) = \beta_0 + \beta_1 * Sibling_d + \beta' \boldsymbol{X}_d + \epsilon_d$$

where d identifies the dyad, Δ is the dyad-specific difference in age-at-death (logged), Sibling is a dummy taking value 1 if dyad d is a sibling pair (0 otherwise), while **X** includes covariates measuring dyad-specific

---

[2] For 0.1% of dyads in our analytical sample, difference in age-at-death measured in months is zero. Since log transformations cannot be applied, these dyads are dropped from the main analysis.

family background characteristics used for the conditional random matching of unrelated individuals. Given such common set of family background characteristics, the coefficient of interest, β_1, estimates sibling similarity in age-at-death as compared to that of two (conditionally) randomly matched unrelated individuals.

4.2 Dyads construction

We construct sibling and unrelated dyads adapting the analytical strategy developed by Raab et al. (2014). We generate sibling dyads by matching each individual in the sample with his sibling(s). For individuals with only one sibling in the sample, there is only possible match. For individuals with two or more siblings in the sample, multiple matches are possible. When running regressions, we account for multiple sibling dyads belonging to the same household of origin by clustering standard errors at the household level. To make sure only unique sibling dyads are considered, we identify doublets, i.e., matches containing the same siblings in different dyadic order, and keep only one match out of each doublet at random. We end up with 174,878 sibling dyads. To generate unrelated dyads, we first match each individual in the sample to all individuals sharing the same family background characteristics, as measured by mother birth cohort, county of residence, urban/rural status, and father occupation in the 1920 Census. We exclude dyads so generated where individuals are, in fact, siblings, and treat doublets as above. Out of all possible matches, we then keep one randomly chosen match per individual. This procedure yields 282,187 unrelated dyads.

5 Preliminary Results

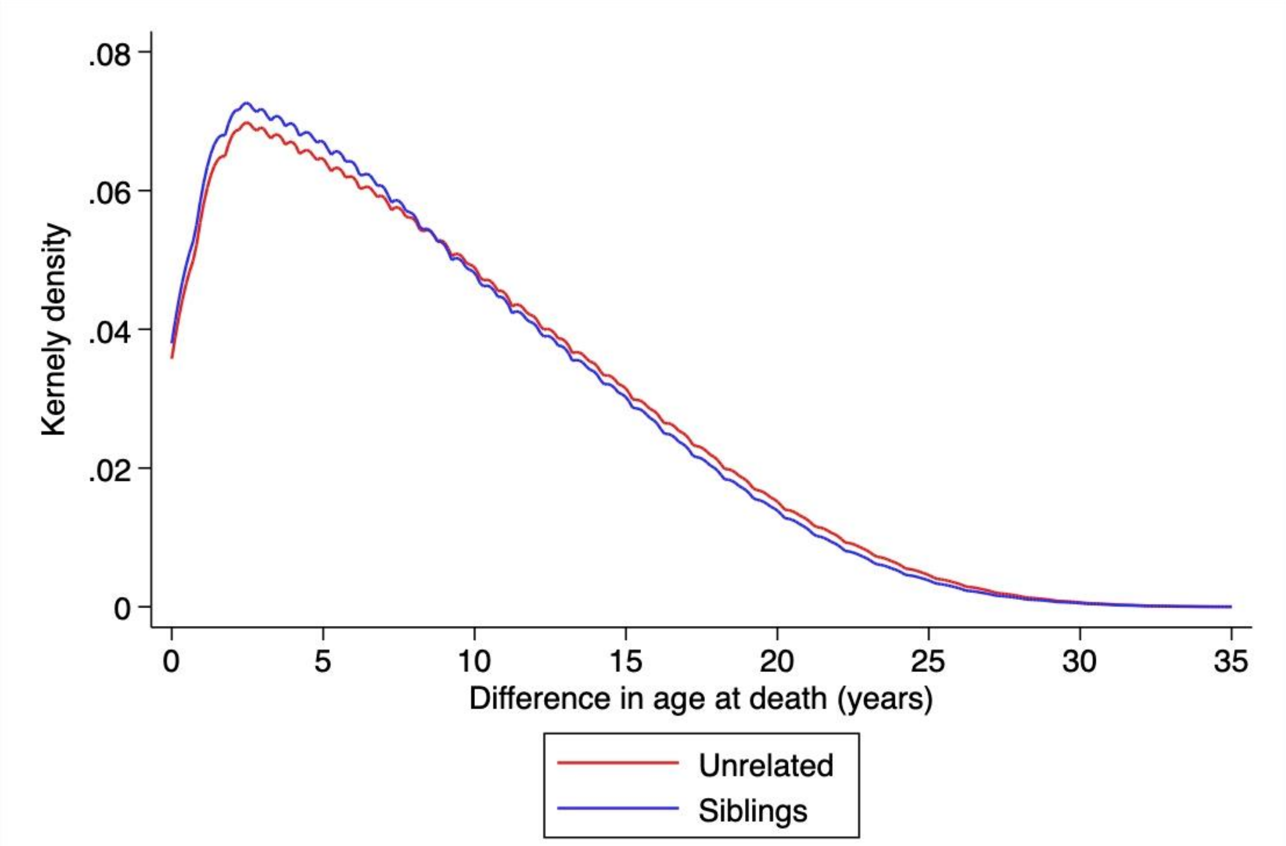
Figure 1 presents results for the difference in age at death in years. Sibling dyads, indicated in blue, tend to die within ten years of each other more than unrelated dyads. After that, there is a crossover, suggesting that at the population level siblings are more similar in lifespan when compared with a control group of unrelated dyads that are constructed to be identical with regard to mother birth cohort, county of residence, urban/rural status, and father occupation in the 1920 Census. This constitute a first finding in support of H1a, which will be further elaborate upon with heterogeneities in socio-economic background, geography, and time.

Figures and Tables

Table 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

Figure 1 – Difference in Age at Death in Log Years between Siblings and Unrelated Dyads



Note. Unrelated individuals conditionally matched based on state and county of residence, urban status and father occupation in 1920. N=457065

References

Abramitzky, R., Boustan, L. and Rashid, M. (2020), Census Linking Project: Version 1.0 [dataset]. https://censuslinkingproject.org.

Alter, G., Oris, M. and Broström, G. (2001), 'The Family and Mortality: a Case Study from Rural Belgium', *Annales de démographie historique* 1(101), 11–31.

Curtis, S. L., Diamond, I. and 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., and Yildirim, U. (2021), CenSoc Mortality File: Version 2.0. Berkeley: University of California.

Guo, G. (1993), 'Use of sibling data to estimate family mortality effects in Guatemala', *Demography* 30(1), 15–32.

Karhula, A., Erolaa, J., Raabb, M. and Fasang, A. (2019), 'Destination as a process', *Advances in Life Course Research* 40, 85–98.

Kröger, H., Hoffmann, R., Tarkiainen, L. and Martikainen, P. (2017), '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.

Kröger, H., Hoffmann1, R., Tarkiainen, L. and 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.

Mandel, M., de Una-Alvarez, J., Simon, D. K. and Betensky, R. A. (2018), 'Inverse Probability Weighted Cox Regression for Doubly Truncated Data', *Biometrics* 74(2), 481–487.

Perls, T. T., Wilmoth, J., Levenson, R., Drinkwater, M., Cohen, M., Bogan, H., Joyce, E., Brewster, S., Kunkel, L. and Puca, A. (2002), 'Life-long sustained mortality advantage of siblings of centenarians', *Proceedings of the National Academy of Sciences* 99(12), 8442–8447.

Raab, M., Fasang, A. E., Karhula, A. and Erola, J. (2014), 'Sibling Similarity in Family Formation', *Demography* 51(6), 2127–2154.

Rennert, L. and Xie, S. X. (2018), 'Cox regression model with doubly truncated data', *Biometrics* 74(2), 725–733.

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

Sasson, I. (2016), 'Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010', *Demography* 53, 269–293.

Scalone, F., Agati, P., Angeli, A. and 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.

van Dijk, I. K. (2019), 'Early-life mortality clustering in families: A literature review', *Population Studies* 73(1), 79–99.