Historical Patterns in the Intergenerational Transmission of Lifespan and Longevity: A Research Note on U.S. Cohorts Born Between 1700 and 1900

Demography (2024)
DOI 10.1215/00703370-11458359 © 2024 The Authors

ADVANCE PUBLICATION

Historical Patterns in the Intergenerational Transmission of Lifespan and Longevity: A Research Note on U.S. Cohorts Born Between 1700 and 1900

Saverio Minardi, Giulia Corti, and Nicola Barban

ABSTRACT This research note examines historical trends in lifespan inequality and the intergenerational transmission of lifespan and longevity in the United States over the eighteenth, nineteenth, and twentieth centuries. We contribute to the literature by expanding the estimates of the familial component beyond parent–child associations to include multigenerational and horizontal classes of relatives of different sexes. We also examine how lifespan inequality and the role of the family in lifespan and longevity changed over time. We address the challenge of studying extended family networks in historical times by leveraging recent online crowdsourced genealogical data. Results confirm the presence of a familial component for all classes of relatives considered and highlight a stronger association for horizontal than for vertical relationships. Despite decreasing lifespan inequality, we find no evidence of decreased familial lifespan stratification throughout history. If anything, the results suggest a strengthening of the parent–child association. Finally, the results contribute to the debate on the representativeness and usability of crowdsourced genealogical data by emphasizing the importance of sample selection based on the quality of the information collected.

KEYWORDS Intergenerational transmission • Longevity • Online genealogies • Multigenerational • Social mobility

Introduction

Throughout history, living conditions in industrialized countries have greatly improved, leading to significant increases in life expectancy and dramatic decreases in lifespan inequalities (Edwards and Tuljapurkar 2005; van Raalte et al. 2018). However, assessing a society's equality level requires consideration of ascriptive characteristics (e.g., family background) to achieve desirable social outcomes, including the possibility of living a long and healthy life. Sociodemographic research has extensively examined the intergenerational transmission (IGT) of health, highlighting the importance of family attributes in shaping longevity and lifespan.[1]

[1] The terms *lifespan* and *longevity* are often used interchangeably. To avoid confusion, we will use *longevity* to refer to survival into old ages and *lifespan* to refer to age at death (van den Berg et al. 2017).

ELECTRONIC SUPPLEMENTARY MATERIAL The online version of this article (https://doi.org/10.1215/00703370-11458359) contains supplementary material.

Despite evidence of a familial component in human longevity (see Table A33; all tables and figures designated with an "A" are available in the online appendix), two areas remain understudied because of limited data availability. First, estimates of IGT of lifespan and longevity are typically limited to parent–child correlations, overlooking the role of extended kinship networks and potentially underestimating the importance of ascriptive characteristics. Broader sets of relatives account for a wider range of social, cultural, and environmental factors that contribute to lifespan stratification (Hällsten 2014; Mare 2011). Second, studies on lifespan correlation have not fully examined the contextual and historical variations in this relationship. Although a historical decline in overall lifespan inequality has been documented (e.g., Vaupel et al. 2011), it is not clear whether IGT in human lifespan was stable or changed.

This study investigates the familial stratification of adult lifespan and longevity and its historical variation in the United States for cohorts born in the eighteenth and nineteenth centuries. First, it expands estimates of the familial component beyond the parent–child relation to include multiple generations of *horizontal* and *vertical* relatives of different sexes. Second, it investigates long-term historical trends in the IGT of lifespan and longevity.

This type of study requires large-scale, multigenerational data on family biographies spanning hundreds of years. Historical sources linking multiple generations are extremely rare, and studies have typically focused on small and selected populations. To overcome these limitations, we used data from the Familinx project, arguably the largest scientific resource of internet-based crowdsourced genealogical microdata covering complex familial relationships over historical times (Kaplanis et al. 2018).

The availability of crowdsourced genealogies raises questions regarding opportunities to use these new data sources in population studies. We contribute to this recent debate (e.g., see Black et al. 2023; Blanc 2023, 2024; Calderón Bernal et al. 2023; Chong et al. 2022; Fire and Elovici 2015; Stelter and Alburez-Gutierrez 2022) by investigating and discussing the biases and representativeness of online-based genealogies and by describing their potential for the study of IGT of demographic behavior over history.

First, we compare estimates from Familinx with more conventional data sources and highlight their historical-demographic consistency. Second, we examine potential sources of bias and reflect on how they affect our findings. Research has addressed *structural biases* in genealogical data that affect family tree completeness and record omission, such as the underrepresentation of childless individuals, premature deaths, and record selection by descent survivorship (Calderón Bernal et al. 2023; Zhao 2001). We additionally describe the presence of *accuracy bias* concerning the level of detail in the reported information for the included records.

Our exploratory analysis revealed a temporal increase in the detail of recorded information, with higher quality records observed among more longevous and likely selected individuals. This trend poses problems of both measurement and selection bias. Noise and error measures might result in attenuation bias, especially in older periods. However, sample restrictions based on information accuracy might result in selected samples. Therefore, we emphasize that the quality of recorded information is not random and suggest that researchers should carefully consider their sample restrictions.

Data and Methods

Familinx data include millions of demographic records and family trees based on crowdsourced genealogies.[2] Our baseline sample includes U.S.-born individuals between 1690 and 1910, with complete information on birth year, death year, sex, and birth location,[3] and complete information on the same variables for both parents. We restrict the sample to individuals and parents who lived to at least age 30 but not beyond age 110. The lower limit is necessary because Familinx data are not acceptably representative of individuals below age 30 (Chong et al. 2022)[4]; the upper limit excludes outliers and unreasonable values.

We exclude individuals born before their parents' 10th birthday, after the father's 65th birthday, after the mother's 55th birthday, or after their mother's death or more than a year after their father's death. Finally, we exclude individuals with more than 20 siblings. Our final sample comprises 350,884 individuals (Figure A2).

In extensive data validation studies, Kaplanis et al. (2018) and (for Europe) Blanc (2023, 2024) found that Familinx data are consistent with official statistics. Nevertheless, we provide further comparisons for the United States in the online appendix (see Data Comparison section). We compare trends in state of birth, sex, and second-generation migrant distributions with the 1850–1940 U.S. censuses. We compare period life expectancy with estimates from the Human Mortality Database, U.S. vital statistics (Linder and Grove 1943), and Hacker (2010). Overall, the Familinx estimates are consistent with the census data. We observe an underestimation of female records and thus perform all analyses separately by sex and parental lineages. Finally, as already documented (Chong et al. 2022; Stelter and Alburez-Gutierrez 2022), genealogical data tend to underestimate mortality levels, especially in older periods, but are fairly consistent in documenting long-term trends.

Our main specification relates the individual's lifespan or longevity to a relative's or group of relatives' lifespan or longevity:

$$L_i = \beta_0 + \beta_1 LR_i + \sum_j a_j \mathbf{X}_{ji} + \varepsilon_i. \quad (1)$$

The first outcome variable (L_i) is the lifespan of an individual i, defined as the individual's deviation in age at death from the cohort-country-sex average.[5] The second dependent variable is longevity, a dichotomous variable indicating whether individuals lived up to the 80th percentile of their cohort-country-sex distribution.

We estimate different models on the basis of the relatives under consideration. The main independent variable (LR_i) is the relative's lifespan or the average lifespan of the group of relatives considered: parents, siblings, grandparents, and cousins.

[2] Data are available at https://osf.io/fd25c/.

[3] Birth location in the Familinx data is provided as unstructured strings and as latitude and longitude. We obtain categorical information on country and state of birth through text similarity algorithms and automated reversed geoparsing (see the online appendix).

[4] As a robustness check, we relaxed this limit and test all relationships using different survival thresholds from ages 0 to 80 (see Tables A27–A32 and Figures A11–A14).

[5] We also tested our specifications using non–de-meaned lifespan and log of lifespan and reached the same conclusions (Tables A27–A32).

We calculate the parental lifespan separately for fathers and mothers; we also calculate the average lifespan of both parents (mid-parent). Similarly, we calculate longevity separately for each parent and the parents jointly, indicating at least one longevous parent. Grandparents' lifespans and longevity are estimated separately for grandfathers and grandmothers by paternal and maternal lines.

Sibling lifespan is defined as the average lifespan of all siblings, excluding the reference individual, and longevity is computed as having at least one longevous sibling. We compute the average lifespan of cousins for the maternal side (individuals whose mothers were sisters) or paternal side (individuals whose fathers were brothers), excluding the reference individual and their siblings. The sample size varies by the relatives analyzed (Table A1).

We estimate family association in lifespan using ordinary least-squares regression. The coefficient for longevity is estimated using logistic regression. Models are run for pooled samples and, when possible, separately by sex. In all analyses, standard errors are clustered at the highest family relationship.

The variables in the vector $\mathbf{X}_{ji}$ include cohort fixed effects, state-of-birth fixed effects, fathers' cohort fixed effects, mothers' cohort fixed effects, grandparents' cohort fixed effects, sex, birth order, and number of siblings; a dummy variable indicating second-generation migrants; and dummy variables indicating imprecise age, father's imprecise age, and mother's imprecise age (see descriptive statistics in Table A2).[6]

Individuals and relatives with precise ages are defined as those with nonmissing and valid entries in their birth and death months and birth and death dates not indicated as "circa." The underlying idea is that if genealogists reported birth and death months, they would likely have had correct information on the years. Testing this assumption, Figure A3 shows age heaping by birth and death year across centuries for the full sample and the subsample of precise records.

Imprecise age and parents' imprecise age are key indicators, given that with noisy measures of relatives' lifespan and longevity, estimates of the intergenerational association will suffer from attenuation bias. As shown in Figure A5, individuals reporting precise ages are likely a selected subsample of the population characterized by a higher average lifespan, lower lifespan inequality, and lower percentage of females. Indeed, individuals whose birth and death dates are recorded down to the month, even centuries after their death, are likely to be selected on socioeconomic background, location, education level, and cause of death.

Therefore, if the full sample suffers from measurement errors, the subsample might be less representative. We test all specifications on both the full sample and subsamples of individuals and relatives reporting the precise ages. Furthermore, we include dummy variables for imprecise age as control variables. Nevertheless, our preferred specifications are based on the full sample because it is larger and more representative of the full population.

To estimate historical trends in IGT, we run the models for parent–child and siblings separately for 200 21-year cohorts from 1690 to 1910. To avoid strict separation between cohorts and maximize sample size, we adopt moving cohorts. We estimate each model on a sample defined as the reference year from 1700 to 1900 plus and minus 10 years—that is, 1700 (1690–1710), 1701 (1691–1711), . . . , 1900 (1890–1910).

[6] We test all specifications with and without controls; all results are reported in the online appendix.

The issue of measurement error is particularly relevant for long-term trends in IGT, given that age misreporting is more common in older periods and progressively declines (see Figure A4). Because the attenuation bias leads to underestimating the coefficient, the progressive decline in measurement error could mechanically lead to an upward trend in IGT. Again, we address this issue by estimating all specifications for the full sample and the selected subsample of the precise dates.

Moreover, IGT trends might be influenced by unrepresentative samples, particularly for non-White populations. Familinx covers primarily Western Europe and North America, suggesting a predominantly European descent among the recorded individuals (Kaplanis et al. 2018). Additionally, the representation of non-White individuals might have increased after the U.S. Civil War. This challenge, which Ward (2023) highlighted, is common in historical studies investigating long-term variations in intergenerational mobility. Estimates based on predominantly White samples might overestimate mobility, especially in older periods, because they do not account for the limited opportunities available for non-White families. Unfortunately, the data do not contain information on race, making it difficult to confront this limitation. However, continuous enhancements in the quality and quantity of genealogical data can address these concerns and racial biases in existing mobility studies.

Results

Figure 1 shows an increase in the average and median lifespan for men and especially women over almost the entire period. Moreover, we observe a general decline in lifespan inequality and a progressive increase in the share of longevous people during the same period.

Figure 1 highlights a lifespan decline for the birth cohorts preceding the U.S. Civil War. Other studies, most of which used smaller samples of selected populations, documented that mortality increased and average height declined in the antebellum years, despite an improvement in economic conditions. This "antebellum puzzle" has generally been explained by limited food availability and disease spread (Fogel 1986; Haines et al. 2003; Pope 1992).

The findings in Figure 1 indicate a correlation between trends in lifespan, longevity, and lifespan inequalities. These trends document the process of mortality compression and rectangularization of human survival curves occurring over this period. As the average lifespan increases and deaths become more concentrated at higher ages, without concomitant increases in the maximum age at death, the dispersion in age at death necessarily decreases (Fries 1980; Wilmoth and Horiuchi 1999).

In addition, Figure 1 demonstrates a reversal in sex differences during the nineteenth century, documented by a few other studies (Beltrán-Sánchez et al. 2014; Fire and Elovici 2015; Martin 1951; McNay et al. 2005; Pope 1992). One potential explanation for this historical reversal is a reduction in childbearing deaths. The female distributions in age at death (Figure A1) show a higher share of deaths at childbearing age in earlier periods than in more recent centuries. Moreover, during the same period, women experienced notable increases in social, political, and economic rights, improving their educational level and socioeconomic position, which likely contributed to their improved lifespan.

Fig. 1 Trends in lifespan, lifespan inequality, and longevity by sex conditional on survival to age 30 (United States, 1690–1910 cohorts)

Panel a of Figure 2 shows the effect of parental lifespan on children's lifespans. Coefficients are reported by child and parent sex. The highest coefficient is that of the mid-parent, whereas the coefficients for single parents are approximately half that of the mid-parent. Our findings align with the idea that each parent's lifespan provides independent data and is subject to measurement errors. We observe no notable disparities between the sexes.

Overall, our results are comparable with previous studies (see Table A33), suggesting that online crowdsourced genealogies are a suitable data source for estimating historical patterns in the IGT of lifespan. Note that many previous estimates are also based on genealogical data and often on selected populations. Despite the limitations of the Familinx data, the sample size and multigenerational characteristics represent a considerable improvement over previous data sources.

In the full set of results from the regression model, we observe that individuals reporting imprecise ages live significantly shorter lives than those with complete information (Table A5). The positive parent–child association is also visible when considering parent–child longevity (Figure 1, panel b).

Panel a of Figure 3 shows the siblings' association in lifespan by sex. The siblings' association highly exceeded the parent–child association and is much stronger for same-sex siblings. Again, longevity associations are strong and significant and show patterns comparable to those for lifespan (Figure 3, panel b).

Figure 4 shows the association with grandparents' lifespan by lineage and grandparent sex, both direct (controlling for parents' lifespan) and total effects.[7] We observe

[7] In estimating direct effects, we control only for parents' lifespan as a mediator of grandparents and cousins; we do not control for siblings' lifespan because having both parents' information was a necessary sample precondition, but not all individuals have a sibling. The inclusion of a siblings' effect further restricts the sample. Nevertheless, we estimate the same model including siblings' lifespan (Tables A13.2 and A17.2); the coefficients are slightly lower, but the patterns and conclusions hold.

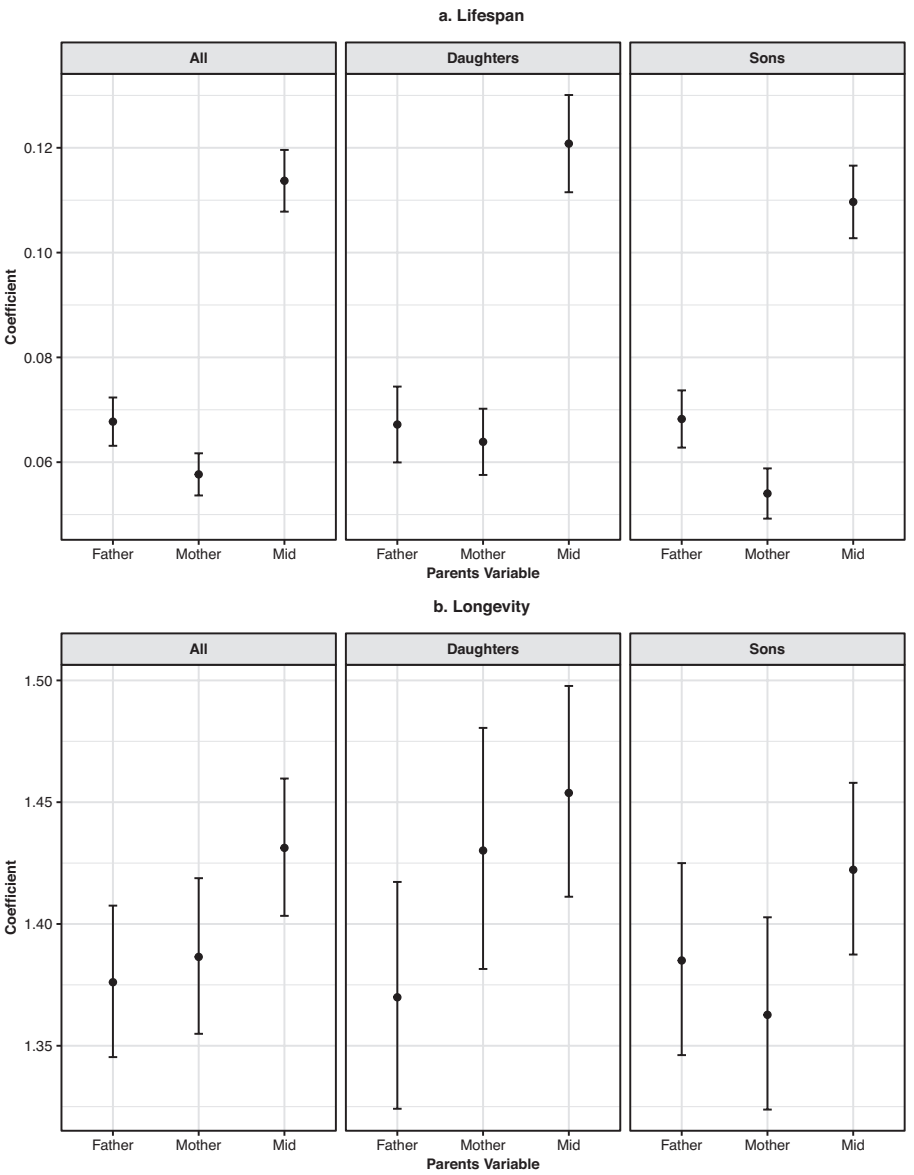

Fig. 2 Intergenerational lifespan (panel a) and longevity (panel b) associations between parents and children by children and parent sex. Panel a reports effects estimated by ordinary least-squares regression of children's lifespan on parents' lifespan for the pooled U.S. sample (1690–1910). Panel b reports effects estimated by binomial logistic regression of children's longevity on parents' longevity for the pooled U.S. sample (1690–1910). Controls include sex; the child's, father's, and mother's birth year; the child's birth state, birth order, and number of siblings; dummy variables indicating the imprecise age at death for the child, mother, and father; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the family level. Full results are shown in Tables A5 (panel a) and A21 (panel b).

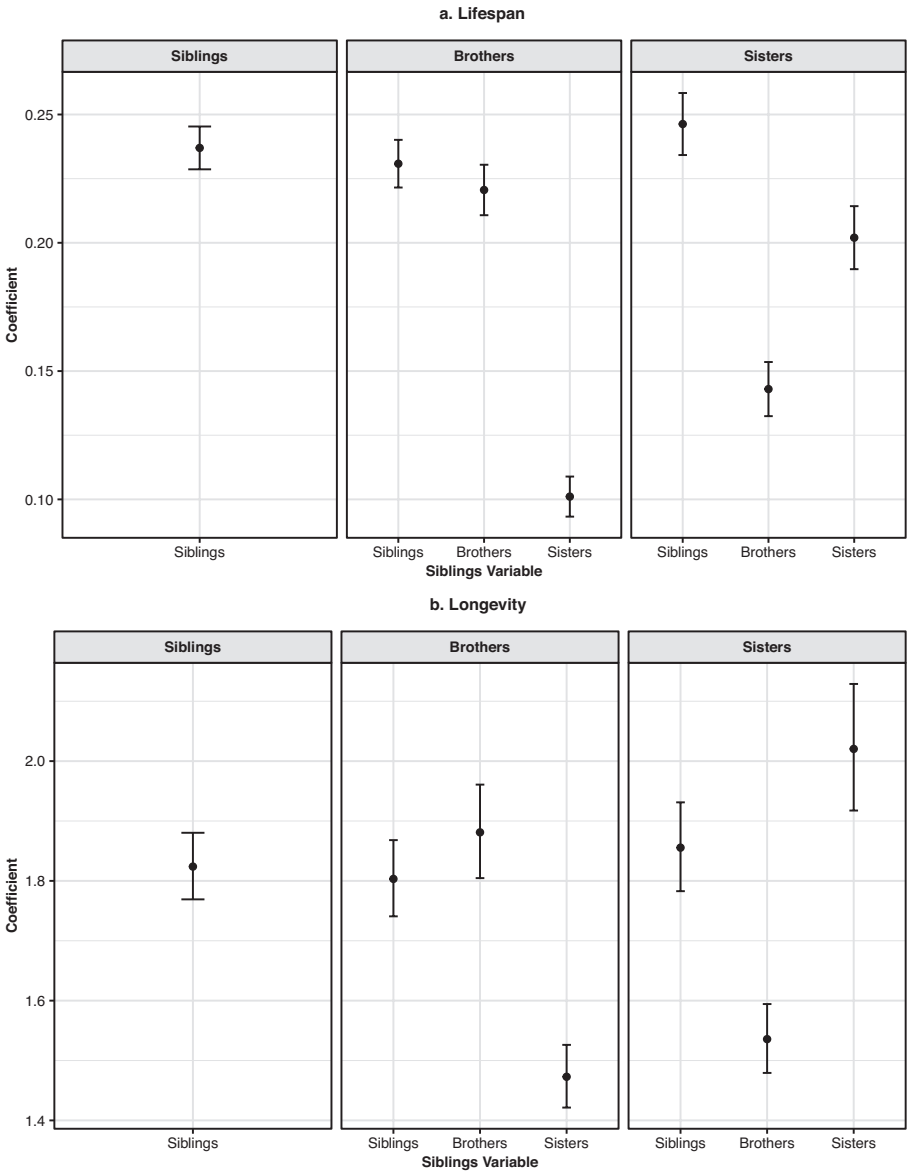

Fig. 3 Association of lifespan (panel a) and longevity (panel b) between siblings by sex. Panel a reports effects estimated by ordinary least-squares regression of the individual's lifespan on siblings' average lifespan for the pooled U.S. sample (1690–1910). Panel b reports effects estimated by binomial logistic regression of the individual's longevity on siblings' longevity for the pooled U.S. sample (1690–1910). Controls include the child's sex, birth year, birth state, birth order, and number of siblings; a dummy variable indicating the child's imprecise age at death; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the family level. Full results are shown in Tables A9 and A25.

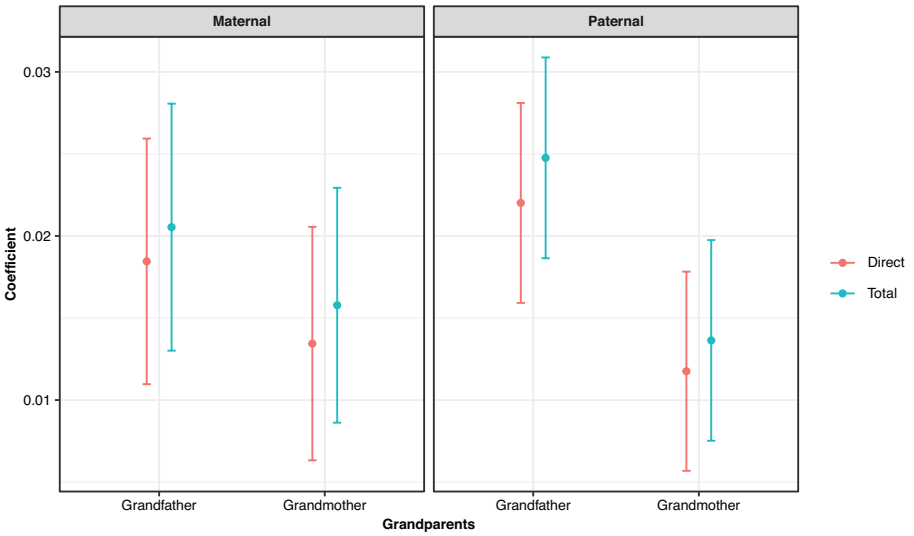

Fig. 4 Lifespan association between grandparents and grandchildren. Effects are estimated by ordinary least-squares regression for the pooled U.S. sample (1690–1910). Direct effect estimates include controls for parents' lifespan. Controls include sex; the child's, father's, mother's, and grandparent's birth year; the child's birth state, birth order, and number of siblings; dummy variables indicating the imprecise age at death for the child, mother, and father; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the grandparent level. Full results are shown in Table A13.

a small but significant association. The total and direct effects were quite similar, suggesting that grandparents have an effect beyond their transmission to parents.

Finally, we examine the cousins' multigenerational and horizontal relationships (Figure 5). These relationships are strong and comparable to parent–child relationships. Similar to siblings, cousins share several generational and environmental conditions at the local level, in addition to broad family characteristics.

Transmission of Lifespan and Longevity Across History

Figure 6 shows the evolution of the parent–child association over centuries. Each point is the estimated coefficient and the shaded areas indicate confidence intervals of a regression of mid-parent lifespan on children's lifespans for birth cohorts defined by the 20 years surrounding the reference year.

The parent–child association shows an upward trend, indicating declining mobility. This result aligns with recent research on the historical evolution of socioeconomic mobility in the United States (e.g., Song et al. 2020).

There are several substantive reasons why intergenerational associations in lifespan may show upward trends. However, the variation in data quality over time is crucial. The proportion of imprecise records is much higher in earlier cohorts and progressively declines over time, possibly causing attenuation bias in the estimated coefficients of earlier periods. To address this limitation, Figure A6 shows the estimation of the same set of models on the subsample of individuals with a precise age

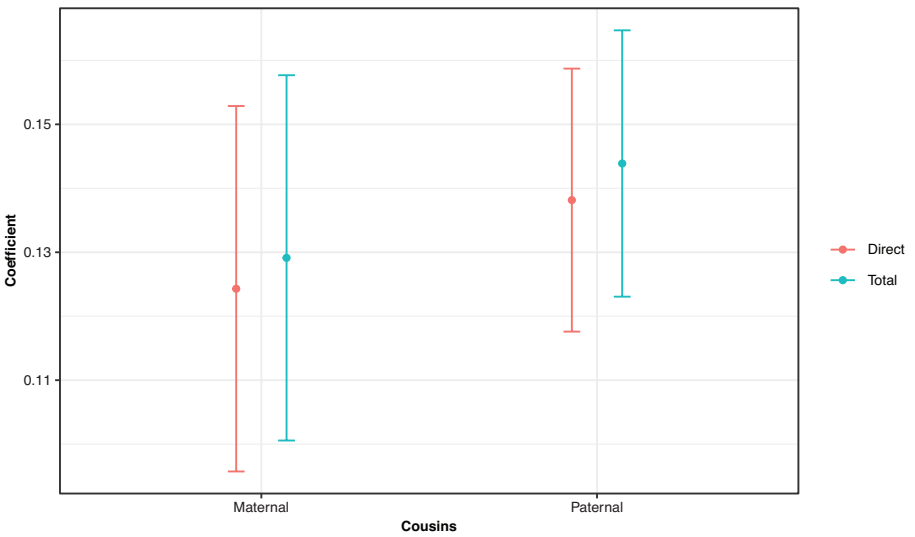

Fig. 5 Lifespan association between cousins by lineage. Effects are estimated by ordinary least-squares regression of the individual's lifespan on cousins' average lifespan for the pooled U.S. sample (1690–1910). Direct effects include parental lifespan. Controls include the child's sex, birth year, birth state, birth order, and number of siblings; a dummy variable indicating the child's imprecise age at death; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the grandparent level. Full results are shown in Table A17.

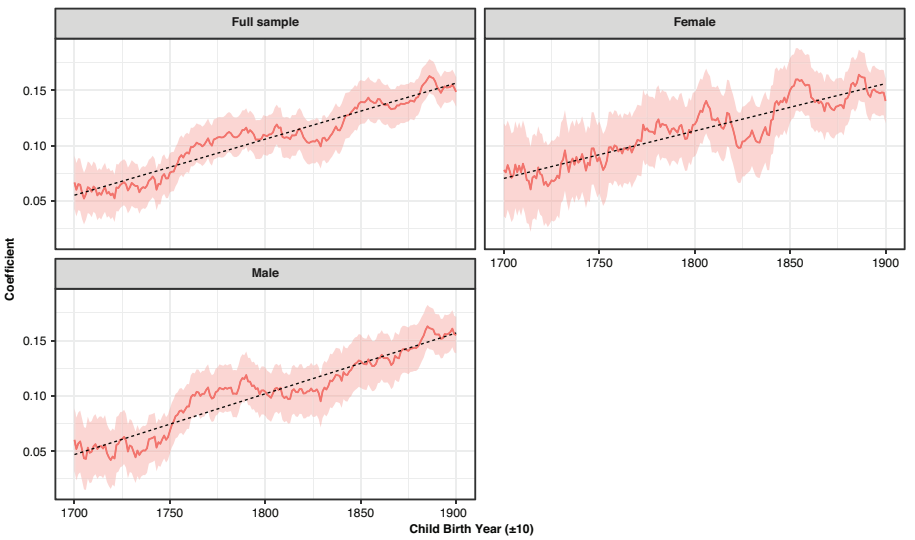

Fig. 6 Association between children's lifespan and mid-parent lifespan by sex and cohort (1700–1900). Effects are estimated by ordinary least-squares regression of children's lifespan on parents' lifespan for separate samples based on a 20-year span from each reference year [1700 (1690–1710), 1701 (1691–1711), . . . , 1900 (1890–1910)]. Controls include sex; the child's, father's, and mother's birth year; the child's birth state, birth order, and number of siblings; dummy variables indicating the imprecise age at death for the child, mother, and father; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the family level.

record for 40-year birth cohorts, given that the sample sizes are smaller in earlier periods. In this case, the coefficient for earlier periods is higher and less precise, resulting in a less drastic but still upward trend. Thus, evidence of a positive IGT trend is present regardless of the sample considered but is less evident when looking at individuals with precise ages at death. However, a subsample based on record accuracy is unlikely to be random, particularly in older periods (Figure A5).

Well-documented genealogies are typically from selected families that report higher average lifespans. We would expect higher lifespans but also higher intergenerational correlations for selected population subgroups, particularly in older periods, resulting in more moderate upward trends, as shown in Figure A6. A less visible upward trend in IGT is because privileged social groups have lived in favorable conditions for centuries, reducing the influence of environmental and social factors, such as poverty and health care availability, on their lifespan. Social deprivation and harsher conditions have had less influence on their survival chances, and improvements in these conditions should not affect the degree of IGT. This characteristic is considered an incidental advantage for biodemographic studies because it minimizes the heterogeneity of the study population (Gavrilov et al. 2002).

Both structural and accuracy biases likely produce more selected samples. The further back in time we go, the more strongly selected these samples become. As a result, the IGT coefficient is likely to be overestimated in older periods, underestimating the upward trend in the parent–child association. Therefore, our depiction of the growth in parent–child lifespan associations across centuries is likely conservative.

Moreover, time trends in the IGT of longevity (Figure 7) show an upward trend for both sexes. Longevity is a much less precise indicator of life duration than lifespan and should therefore be less influenced by age misreporting. In this case, errors are relevant as far as they shift individuals below or over the 80th percentile.

The evolution of the sibling association is shown in Figure 8. In contrast to the parent–child association, the sibling association shows no evidence of an upward trend in lifespan or longevity, regardless of whether precise records are included or excluded.

The observed difference in IGT trends estimated through siblings and parents is an intriguing puzzle. One possible explanation is that intergenerational differences in social status, lifestyle, and living conditions gradually decline, but intragenerational differences in these factors do not. Because siblings typically belong to similar generations, their shared environment and lifestyles remain constant over time. For instance, diseases and causes of death might differ more between parents and children in earlier periods than in more modern times. In contrast, siblings experience the same environment regardless of the period considered. Moreover, stronger geographic and occupational mobility in the past might have created physical and social distance between parents and children but not between siblings, who often experienced the same transitions. This idea echoes the Song et al. (2020) suggestion that occupational mobility declines resulted from an intergenerational shift caused by children's movement from farms to manufacturing sectors during industrialization. Another possible explanation is the overestimation of siblings' coefficients in the past. Genealogies tend to emphasize and more accurately document vertical rather than horizontal relationships. As shown in Figure A15, fewer siblings were documented in earlier times. Those who were recorded were likely selected and lived longer, artificially inflating

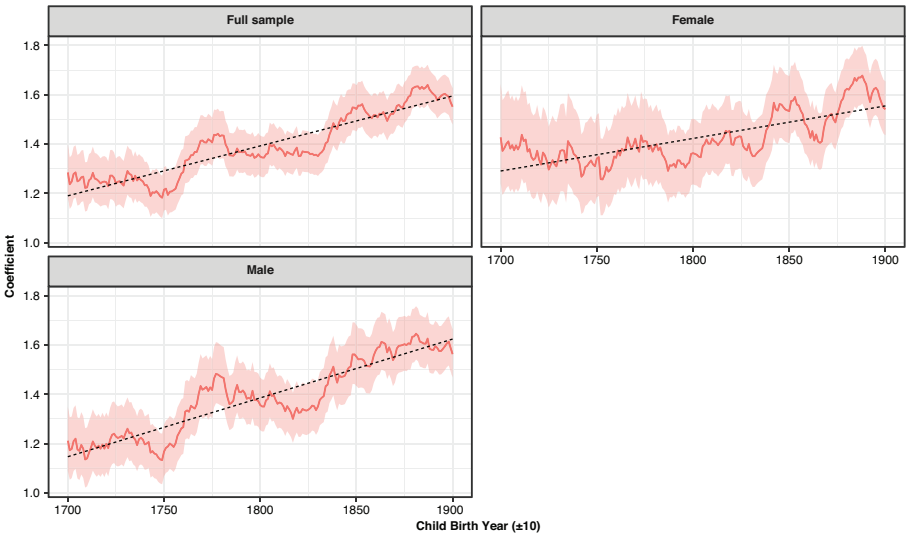

Fig. 7 Association between children's longevity and mid-parent longevity by sex and cohort (1700–1900). Estimates, reported as odds ratios, are estimated through binomial logistic regression of children's longevity on parents' longevity for separate samples based on 20-year intervals around each reference year [1700 (1690–1710), 1701 (1691–1711), . . . , 1900 (1890–1910)]. Controls include sex; the child's, father's, and mother's birth year; the child's birth state, birth order, and number of siblings; dummy variables indicating the imprecise age at death for the child, mother, and father; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the family level.

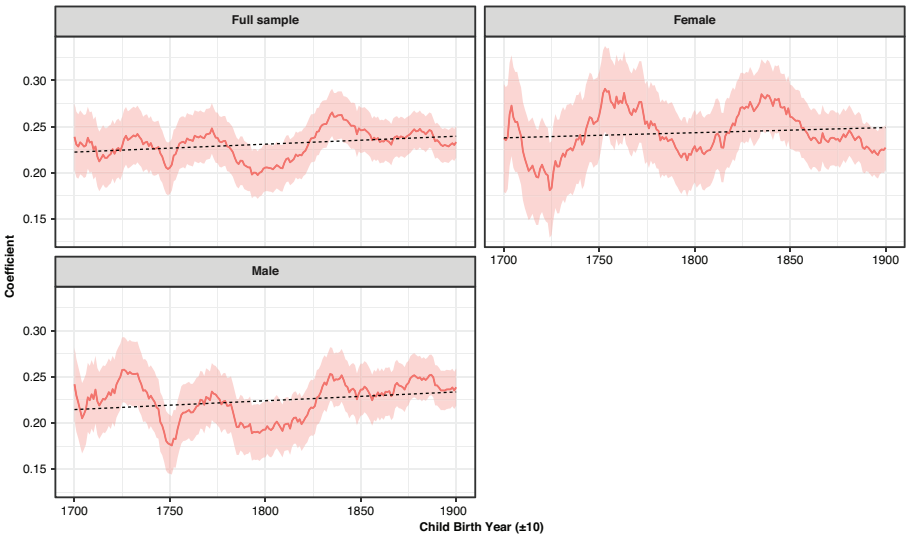

Fig. 8 Siblings' association in lifespan by birth cohort and sex. Effects are estimated by ordinary least-squares regression of lifespan on average sibling lifespan for separate samples based on a 20-year span from each reference year [1700 (1690–1710), 1701 (1691–1711), . . . , 1900 (1890–1910)]. Controls include the child's sex, birth year, birth state, birth order, and number of siblings; a dummy variable indicating the child's imprecise age at death; and a dummy variable indicating second-generation migrant. Standard errors are clustered at the family level.

the coefficients in earlier periods and hindering the upward trend observed for parent–child associations.

The preceding results are estimated for individuals who survived for at least 30 years. A lower bound is a common practice in the study of intergenerational lifespan correlation (see Table A33), which also reflects that parents' lifespan is necessarily left-truncated. Nevertheless, we also tested our findings using different age thresholds from 0 to 80 (Tables A27–A32 and Figures A11–A14). The general conclusions remain unchanged. In line with previous studies, higher thresholds resulted in stronger associations, suggesting that the IGT of longevity is stronger than that of lifespan (Gavrilov and Gavrilova 2001; van den Berg et al. 2017). Lower age thresholds also produced higher average coefficients, consistent with previous studies, suggesting a strong familial clustering of child and infant mortality (van Dijk 2019; van Dijk and Mandemakers 2018). Moreover, at younger ages, the Pearson correlation coefficients were lower than the regression coefficients. Parents' distributions have a much lower variance than probands because we are not conditioning on children reaching adulthood but are necessarily doing so for parents (Piraino et al. 2014). Nevertheless, the results for lower thresholds should be viewed cautiously, given the well-documented underrepresentation of early deaths.

Conclusions

Our results confirmed the existence of vertical association in lifespan. We also highlighted the presence of multigenerational and horizontal associations and showed that the strongest familial component is evident in horizontal relationships.

We observed a long-term increase in the average lifespan and a concomitant decrease in lifespan inequality. These processes resonate with a general trend toward a rectangularization of survival curves and mortality compression, consistent with previous studies documenting a negative association between longevity and lifespan dispersion in more recent periods (Aburto et al. 2020; Vaupel et al. 2011; Wilmoth and Horiuchi 1999).

However, a decline in general levels of lifespan inequalities was not associated with a decline in IGT. If anything, our results indicated a strengthening of the parent–child association. Sibling associations, on the other hand, remained stable over the study period. These results suggest that decreases in the uncertainty in the ages at death are not necessarily followed by declines in lifespan inequalities based on ascribed characteristics and could even reinforce them.

Limited sample sizes complicate the measurement of long-term changes in kin beyond parents and siblings, especially in older periods, and result in noisy and unreliable historical estimates for these relationships. Despite these difficulties, studying the variation in the association with distant relatives is an intriguing area of inquiry and could provide insights into the evolution of family structures from extended to nuclear groups. As the quantity and quality of records continue to improve, the study of historical variation in lifespan association between more distant relatives presents a promising avenue for future research.

Although genealogies have frequently been used for studying the IGT of lifespan and longevity (Gavrilov et al. 2002), they are not without limitations (Hollingsworth

1976). One bias arises from the fact that the inclusion of individuals in a family tree is often contingent on the existence of living descendants. This selection might lead to structural biases, that is, biases based on the omission of family trees or records (Zhao 2001).

Another, less-discussed source of bias is accuracy bias, which concerns the details of the reported information for the included records. On the one hand, measurement error might result in attenuation bias and the overestimation of the upward IGT trend. On the other, we highlighted that individuals with detailed information lived longer, on average. This finding suggests that individuals for whom high-quality information is retrievable might be a selected subgroup of more educated or higher class individuals, especially in older periods, supporting the idea that sample restrictions based on records accuracy might result in an underestimation of the upward IGT trend.

Another source of bias to consider is the probable underestimation of non-White individuals, potentially leading to an overestimation of mobility because the limited opportunities available to non-White families are overlooked. Unfortunately, the data lack a race indicator, preventing us from addressing this issue. Nonetheless, with the continuous improvement in record accuracy and volume, genealogy data can help address race bias in historical intergenerational mobility studies. ■

Acknowledgments This project received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement 865356).

References

Aburto, J. M., Villavicencio, F., Basellini, U., Kjærgaard, S., & Vaupel, J. W. (2020). Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences, 117,* 5250–5259.

Beltrán-Sánchez, H., Finch, C. E., & Crimmins, E. M. (2014). Twentieth century surge of excess adult male mortality. *Proceedings of the National Academy of Sciences, 112,* 8993–8998.

Black, S. E., Duzett, N., Lleras-Muney, A., Pope, N. G., & Price, J. (2023). *Intergenerational transmission of lifespan in the U.S.* (NBER Working Paper 31034). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w31034

Blanc, G. (2023). *Demographic transitions, rural flight, and intergenerational persistence: Evidence from crowdsourced genealogies* (Report). Manchester, UK: University of Manchester. Retrieved from https://www.guillaumeblanc.com/files/theme/Blanc_crowdsourced.pdf

Blanc, G. (2024). *The cultural origins of the demographic transition in France* (Lewis Lab Working Paper 2024-02). Manchester, UK: University of Manchester. Retrieved from https://www.guillaumeblanc.com/files/theme/Blanc_secularization.pdf

Calderón Bernal, L. P., Alburez-Gutierrez, D., & Zagheni, E. (2023). *Analysing biases in genealogies using demographic microsimulation* (MPIDR Working Paper, No. WP-2023-034). Rostock, Germany: Max Planck Institute for Demographic Research. https://doi.org/10.4054/mpidr-wp-2023-034

Chong, M., Alburez-Gutierrez, D., Del Fava, E., Alexander, M., & Zagheni, E. (2022). *Identifying and correcting bias in big crowd-sourced online genealogies* (MPIDR Working Paper, No. WP 2022-005). Rostock, Germany: Max Planck Institute for Demographic Research. https://doi.org/10.4054/mpidr-wp-2022-005

Edwards, R. D., & Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review, 31,* 645–674.

Fire, M., & Elovici, Y. (2015). Data mining of online genealogy datasets for revealing lifespan patterns in human population. *ACM Transactions on Intelligent Systems and Technology, 6,* 28. https://doi.org/10.1145/2700464

Fogel, R. W. (1986). Nutrition and the decline in mortality since 1700: Some preliminary findings. In S. L. Engerman & R. E. Gallman (Eds), *Long-term factors in American economic growth* (pp. 439–556). Chicago, IL: University of Chicago Press.

Fries, J. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine, 303,* 130–135.

Gavrilov, L., & Gavrilova, N. (2001). Biodemographic study of familial determinants of human longevity. *Population: An English Selection, 13*(1), 197–221.

Gavrilov, L., Gavrilova, N., Olshansky, S. J., & Carnes, B. A. (2002). Genealogical data and the biodemography of human longevity. *Social Biology, 49,* 160–173.

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

Haines, M. R., Craig, L. A., & Weiss, T. (2003). The short and the dead: Nutrition, mortality, and the "Antebellum puzzle" in the United States. *Journal of Economic History, 63,* 382–413.

Hällsten, M. (2014). Inequality across three and four generations in egalitarian Sweden: 1st and 2nd cousin correlations in socio-economic outcomes. *Research in Social Stratification and Mobility, 35,* 19–33.

Hollingsworth, T. H. (1976). Genealogy and historical demography. *Annales de Démographie Historique, 1976*(1), 167–170.

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

Linder, F. E., & Grove, R. D. (1943). *Sixteenth census of the United States: 1940. Vital Statistics Rates in the United States, 1900–1940*. Washington, DC: U.S. Government Printing Office.

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

Martin, W. J. (1951). A comparison of the trends of male and female mortality. *Journal of the Royal Statistical Society: Series A (General), 114,* 287–306.

McNay, K., Humphries, J., & Klasen, S. (2005). Excess female mortality in nineteenth-century England and Wales: A regional analysis. *Social Science History, 29,* 649–681.

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.

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

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,* 251–258.

Stelter, R., & Alburez-Gutierrez, D. (2022). Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics. *Proceedings of the National Academy of Sciences, 119,* e2120455119. https://doi.org/10.1073/pnas.2120455119

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

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

van Dijk, I. K., & Mandemakers, K. (2018). Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833–1912. *Historical Life Course Studies, 7,* 28–46.

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

Vaupel, J. W., Zhang, Z., & van Raalte, A. A. (2011). Life expectancy and disparity: An international comparison of life table data. *BMJ Open, 1,* e000128. https://doi.org/10.1136/bmjopen-2011-000128

Ward, Z. (2023). Intergenerational mobility in American history: Accounting for race and measurement error. *American Economic Review, 113,* 3213–3248.

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

Zhao, Z. (2001). Chinese genealogies as a source for demographic research: A further assessment of their reliability and biases. *Population Studies, 55,* 181–193.

Saverio Minardi (corresponding author)
saverio.minardi2@unibo.it

Minardi • Department of Statistical Sciences, University of Bologna, Bologna, Italy; https://orcid.org/0009-0003-0984-2327

Corti • Centre d'Éstudis Demogràfics, Barcelona, Spain; https://orcid.org/0000-0002-2287-2307

Barban • Department of Statistical Sciences, University of Bologna, Bologna, Italy; https://orcid.org/0000-0002-4362-4652