

RESEARCH ARTICLE

Trends in the Female Longevity Advantage of Nineteenth-Century Birth Cohorts: Exploring the Role of Place and Fertility

Jason Fletcher¹ , Michael Topping² , and Won-tak Joo³

¹La Follette School of Public Affairs, Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, USA, ²Department of Sociology, University of Wisconsin-Madison, Madison, WI, USA and ³Department of Sociology and Criminology & Law, University of Florida, Gainesville, FL, USA
Corresponding author: Michael Topping; Email: mtopping@wisc.edu

(Received 18 July 2023; revised 21 July 2024; accepted 13 March 2025; first published online 07 May 2025)

Abstract

This article uses online genealogy data from the United States over the nineteenth century to estimate period and cohort-based sex differences in longevity. Following previous work, we find a longevity reversal in the mid-nineteenth century that expanded rapidly for at least a half-century. For measures of conditional survival past childbearing age, females enjoyed a longevity advantage for the whole century. Unlike most mortality databases of this period, genealogical data allow analysis of spatial patterns and the impacts of fertility on longevity. Our results suggest very limited evidence of spatial (state) variation in these patterns. We do, however, find evidence that the associations between fertility and longevity partially explain the trends.

Keywords: Longevity advantage; place; fertility; nineteenth century

Background

Sex differences in mortality have been documented for centuries, with an advantage in life expectancy for females existing essentially since the mid-eighteenth century (Kalben, 2000). However, how long this advantage has existed across contexts is relatively unknown as is whether this advantage existed for males in the past. For instance, one article by Bergeron-Boucher et al. (2020) found that the likelihood of males outliving females varies substantially across dozens of countries and regions since the middle of the eighteenth century. Nevertheless, the robustness of this advantage has not been examined thoroughly in countries such as the United States, which lack long historical demographic records.

In countries with robust longevity registries, a persistent female advantage has grown substantially over time. Due to economic development and improved living conditions for females, life expectancy increased considerably throughout much of the twentieth century for women in most Western societies (Rigby and Dorling, 2007). For instance, Oksuzyan and colleagues (2008) found among Nordic countries that the gap between males and females was between 2 and 4 years during the period 1850–1950. This difference increased to 6–7 years between the period 1950–1980, followed by a decrease since then of 4.5–5 years. The gap in longevity has begun to narrow to some degree, due to a variety of factors, such as health behaviors like smoking, or an uptick in cardiovascular disease, cancer, and other chronic conditions (Sundberg et al., 2018; Zafeiris, 2020). Ultimately, in virtually all modern populations, women have seen both increased longevity and lower risks of death, even in extreme mortality contexts such as famines and epidemics (Luy and Gast, 2014; Zarulli et al., 2018).

In addition to measures of baseline life expectancy or mortality, alternative measures of population health have also viewed females favorably. For instance, females experience lower lifespan variation relative to males (Aburto et al., 2020; Colchero et al., 2016). Essentially, this means that males typically see a greater uncertainty in the length of life due to this increased variation. Measures of lifespan variation build on standard life expectancy estimates but reflect disparities between groups at the population level and differential uncertainty at the individual level (Edwards and Tuljapurkar, 2005; Sasson, 2016). For instance, Sasson (2016) looks at trends in lifespan variation between sex, race, and education groups from 1990 to 2010 and finds that lifespan variation has increased among the lowest educated, whereas the higher educated groups saw declines. Nevertheless, while differences in mortality, longevity, and other measures have shown sex disparities across time and place, finding robust results, these trends, specifically related to measures of longevity in the nineteenth century, remain less clear in the United States.

Historical sex differences in life expectancy in the United States have not been formally documented, compared to some other nations. Nationwide data for the United States for mortality only exist from 1933 onward. Despite this limitation, previous research, through the use of innovative methods, has attempted to construct estimates for the United States dating back to the nineteenth century (Hacker, 2010; Pope, 1992). Ultimately, this prior research found that throughout the nineteenth century, males and females largely possessed roughly equal life expectancy, with the former having a slight advantage. For example, Goldin and Lleras-Muney (2019) show that younger-aged females (ages 5–25) in late 1800s Massachusetts were disproportionately affected by infectious disease, which undoubtedly contributed to a male advantage in life expectancy for these age groups. However, this same article also shows that by the end of the nineteenth centuries, male mortality exceeded female mortality, and this advantage continued in subsequent years, during which the female advantage was largely maintained. Other research indicates that birth cohorts averaged across 13 countries from 1840–1859, 1860–1879, and 1880–1899 all see higher male-to-female mortality ratios after age 40, with earlier cohorts having smaller ratios, and the latter having larger (Beltrán-Sánchez et al., 2015), though less is

known for mortality before age 40 and the related question of maternal mortality.

Explanations for sex differences in longevity

The differences in longevity stem from a combination of biological differences and social and behavioral factors (Austad and Fischer, 2016; Crimmins *et al.*, 2019; Rogers *et al.*, 2010). Regarding the former, biological differences in genetic structure and hormonal differences are viewed by some to favor female life expectancy (Austad, 2006). Where males and females have similar lifestyles, women still showcase a longevity advantage, implying a role for a biological advantage (Brønnum-Hansen and Juel, 2001; Lindahl-Jacobsen *et al.*, 2013). Even among other species in the animal kingdom, evidence indicates that females see an advantage relative to males (Seifarth *et al.*, 2012). Another noteworthy explanation from the evolutionary perspective is the grandmother hypothesis, which posits that older women are able to increase fitness and health by caring for grandchildren as allomothers (Blell, 2017).

Behavioral factors may explain the majority of the disparities in longevity (Janssen, 2020; Rogers *et al.*, 2010). First, among males, their higher prevalence of risk-taking behaviors especially during younger years puts them at risk of not surviving at older ages. Second, males on average have a higher consumption of tobacco¹ and alcohol, drive less safely, and eat less healthily than their female counterparts, which in turn heightens their risk for the development of many chronic conditions and accidents (Beltrán-Sánchez *et al.*, 2015).

In the context of the nineteenth century, however, there may be alternate justifications for longevity differences. Early-life infections and subsequent infant and child mortality may be one mechanism for a male advantage. During this period, public health initiatives to help control the burden of infectious disease received little investment. Urbanized areas in the United States, in particular, essentially suffered an “urban mortality penalty” due to the lack of basic sanitation, contributing to a quicker spread of infectious and water-borne diseases (Condran and Crimmins-Gardner, 1978; Haines, 2001). Aside from a geographic advantage during this period, nearly one in five children did not survive to the age of 5, among both advantaged and disadvantaged populations (Preston and Haines, 1991). Moreover, the majority of children who died before the age of 5 were males (Drevenstedt *et al.*, 2008), but between the ages of 5 and 25, infectious disease disproportionately affected females (Goldin and Lleras-Muney, 2019).

Sex differences in life expectancy throughout the century were due to adult mortality rates from specific illnesses declining more quickly among females. Beltrán-Sánchez *et al.* (2015) examined birth cohorts from the nineteenth century and found that male–female mortality ratios in the latter half of the century increased in countries like Italy, France, and Spain, but other countries, such as the United States and Norway, did not experience these trends until later years. They also found that among specific causes of death, increases in cardiovascular mortality

¹Furthermore, in developed nations, the greatest risk factor that contributes to differences in longevity is tobacco use (Preston and Wang 2006).

among those age groups contributed most significantly to these rising mortality ratios. The rise of mortality from chronic illness, driven by social and behavioral factors such as increased tobacco usage among men or difficulties in processing and coping with stress, helps to explain why females towards the end of the nineteenth century acquired a longevity advantage over males (Mosca et al., 2011; Weidner, 2000).

Another explanation for sex differences that has been offered focuses on childbearing: for example, evolutionary biology posits that reproduction often comes at a cost to the human lifespan, meaning that lower parity essentially helps to give rise to longer lifespans (Ehrlich, 2015). These investigations offer mixed findings, with some studies finding positive impacts of childbearing on longevity (McArdle et al., 2006; Muller et al., 2002) whereas others have found negative correlations (Lycett et al., 2000)². Altogether, a fuller understanding of the strength of this hypothesis likely requires much larger samples to ensure greater reliability.

Finally, the role that geography plays in life expectancy has been thoroughly documented in the United States (Chetty et al., 2016; Deryugina and Molitor, 2021). This is due to the influence of factors such as policy, which are place-specific, depending on the nation, state, or city in which one lives (Montez et al., 2020). Aside from policy differences, there also could be socioeconomic differences that vary geographically that in turn affect mortality (Hayward et al., 1997). Furthermore, a great deal of research looks to examine demographic phenomena at smaller levels of geography, such as Boing et al. (2020), who looked specifically at geographic variation of life expectancy in the United States and found that smaller levels of geography explained greater levels of variation. Thus, it is important to consider smaller units of geography, such as those at the state level, to gauge if there is a relationship between place and health.

Taken together, it is critical to consider each of the above approaches in the study of sex differences in life expectancy during the nineteenth century, a period where there was a lack of robust mortality data (Hacker, 2010). We thus extend work that has been limited to cohort-level data or data from a single state (Beltrán-Sánchez et al., 2015; Goldin and Lleras-Muney, 2019). Also, it is important to give attention to measures that relate to childbirth and geography to see the potential influences that played in these differences in life expectancy.

This article seeks to examine how sex differences in longevity in the United States have changed over time, specifically across nineteenth-century birth cohorts. Specifically, this article addresses the following: (1) the sex differences in longevity and the emergence of a female advantage from mid-century onward; (2) whether these differences differ spatially by looking at the impact of the state of birth; and (3) how parity shapes these differences. Ultimately, the aim of this article is to increase our understanding of the dramatic societal change undergone historically with regard to the sex gap in longevity.

²It should be noted that many of the studies that seek to examine this phenomenon often have severe limitations due to their focus on specific populations. For example, in their study, McArdle and colleagues (2006) look at longevity patterns amongst Amish individuals who survived to 50 years of age and have children, with a final sample size of just over 2,000 individuals.

Methods

The data used in this research comes from genealogical data on the website Geni.com (<https://familinx.org>), which stores individual profiles that have been uploaded to family trees (Kaplanis *et al.*, 2018). Specifically, the site automatically analyzes the user profiles to ascertain any similarities and then gives the option to merge matched profiles. Novel data sources have become promising resources for demographers due to their large sample sizes and ability to cover long historical periods (Alburez-Gutierrez *et al.*, 2019) when alternative data (registry, etc.) are unavailable. Initially, this article's authors collected, cleaned, and later validated for use 86 million profiles from the site, with some dating as far back as the seventeenth century and providing mortality data through 2015 (Kaplanis *et al.*, 2018). From this, the authors gathered demographic information from the collected profiles, specifically exact birth and death dates. Ultimately, the data reflected both events and trends in history (i.e., deaths from the American Civil War, the 1918 influenza pandemic, etc.). Not only this, but the lifespan data that Geni provides have been compared with the average lifespan from a worldwide historical analysis (Oeppen and Vaupel, 2002) to gauge its validity. Kaplanis and colleagues (2018) found correlations of $R^2 = 95\%$ between historical data (Oeppen and Vaupel, 2002) and Geni data, and a 98 percent concordance with countries in the Human Mortality Database (HMD).

The Geni data are a unique source of data, given that it provides large individual-level data on longevity and other demographic characteristics in the nineteenth-century United States. Among the initial 86 million individuals in the dataset, we focus on the US-born with valid information about sex ($N = 5,479,789$). We further restrict our sample to those with valid birth and death dates – based on which the key outcome of this article, *longevity*, is ascertained – and born in the nineteenth century ($N = 1,215,253$). Additionally, we drop those who died younger than 5 years considering that the inclusion in the genealogy is highly selective on survival to early childhood (Kaplanis *et al.*, 2018) and those with unreliable lifespan lengths (>110 years) ($N = 1,161,469$) (for documentation on the sample size reductions, see Table SA, Appendix, Supplementary material). Using the demographic information given by the Geni data, we estimate regression of longevity on 10-year *birth cohorts* and *sex*, after controlling for regional differences using dummies for *birth states*.

Fertility is measured by the number of parent–child relationships reported in the genealogy. Since we cannot differentiate childless individuals from those with missing information on parent–child relationships in the Geni dataset, we restrict our analysis of longevity and fertility to those who reported at least one child ($N = 240,907$) and incorporate four dummies for fertility (having two; three; four; or $\geq$ five children). Considering that childbearing requires survival to reproductive ages, analyses with fertility are conditional on the start (age 15) or end (age 50) of the childbearing window.

Despite clear advantages of the Geni dataset, such as a large sample size and fruitful demographic information at an individual level, there are some possible pitfalls to the data source, like all genealogical data. First, the dataset occasionally includes incomplete demographic information, resulting in reduced sample size and

increased risk of confounding bias. For example, as shown in Table SA, we lose 32 percent ($= 1,770,890/5,479,789$) among the US-born due to missing birth date, and 27 percent ($= 441,252/1,770,890$) among the nineteenth-century US-born due to right censoring (i.e., missing in death date). The rate of censoring is – given that it is cumulative throughout life – comparable to that of modern longitudinal studies ranging from 15 percent to 50 percent (Herd, Carr, and Roan, 2014; Mann and Honeycutt, 2016; Schoeni and Wiemers, 2015), but still problematic since we have no direct information to adjust for confounding due to non-random attrition. Second, even when assuming non-missing information, the inclusion in the geology itself is associated with the complex but unobservable selection process, possibly via socioeconomic status or the number of descendants (Stelter and Albrez-Gutierrez, 2022).

We address potential concerns of sample selection using pieces of representative datasets available. First, we opt to show how our data matches up against cohort estimates from the HMD. Since it has been established in prior work (Kaplanis et al., 2018), that the HMD has a 98 percent concordance with the Geni data, we expect our data to follow similar trends. Second, we compare our estimates to those of the United States Social Security Administration to check the amount and direction of bias our analysis might have. Even in these comparisons, we could not directly check the representativeness of our analytic sample due to no comparable population datasets from the nineteenth century. We note that other datasets including similar information laid out in this article are rather limited, due to both the smaller sample size and temporal constraints. This is especially true for the United States since national administrative record keeping for factors such as mortality and longevity only began in the 1930s (Hacker, 2010). The issues of selection bias, along with other potential limitations that will be discussed later, are important to consider moving forward.

The main analyses are presented as follows. First, we present our cohort longevity estimates (conditional on survival to age 5) from our genealogical data. Next, we then display our results that look at lifespan variation (Aburto et al., 2020; Edwards and Tuljapurkar, 2005). Finally, we present our results on fertility by looking at the number of children an individual has, first for the whole sample and then by gender, in models that are conditional on survival to ages 15, and then age 50, to account for the window of childbearing.

Finally, we further explain the role that fertility plays in the trends observed through a Kitagawa–Blinder–Oaxaca decomposition of the differences in longevity before and after 1870, the point at which trends in longevity accelerate among females, by gender. We opted to use this method based on our findings that fertility decreased over time, and having fewer children became more advantageous over time. The equation is as follows:

$$\bar{Y}_A - \bar{Y}_B = \underbrace{\sum_j \beta_{Bj} (\bar{X}_{Aj} - \bar{X}_{Bj})}_{\text{Endowment Effects}} + \underbrace{\sum_j (\beta_{Aj} - \beta_{Bj}) X_{Bj}}_{\text{Coefficient Effects}} + \underbrace{\sum_j (\beta_{Aj} - \beta_{Bj}) (\bar{X}_{Aj} - \bar{X}_{Bj})}_{\text{Interaction Effects}}$$

$\bar{Y}_B$ and $\bar{Y}_A$ denote longevity before and after 1870, whose difference is expressed by regression coefficient β and distribution X of each fertility level j . We conducted the three-fold decomposition (Daymont and Andrisani, 1984) from the perspective of

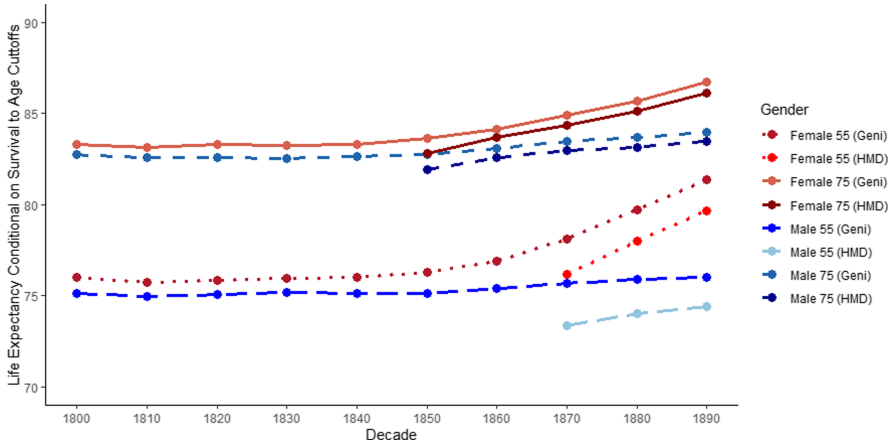


Figure 1. Geni data compared with human mortality database estimates.

older cohorts born before 1870, where we differentiate the endowment (i.e., the effects of decreases in fertility levels $\bar{X}_{Aj} - \bar{X}_{Bj}$, when fixing the coefficients to those of older cohorts β_{Bj}), coefficient (i.e., the effects of changes in coefficients of fertility $\beta_{Aj} - \beta_{Bj}$, when fixing the level of fertility to that of older cohorts X_{Bj}), and interaction effects (i.e., the changes in the level and coefficients of fertility taken together). To examine the distribution and effects of fertility in more detail, we considered 10 fertility dummies, where the last one is for 10 or more children. All the dummy variables (for fertility and birth states) are standardized with zero mean and unit variance and incorporated into the regression models to eliminate the dependence on the choice of the reference category. To limit our attention to only the role of fertility, decomposition was done after controlling for birth state dummies.

Results

Figure 1 displays our cohort estimates from 1800 to 1890 compared to cohort estimates from the HMD. Specifically, we show longevity trends conditional on survival to age 55 and 75, from 1870 and 1850 onwards, respectively. Our data largely mirror the sex-specific trends that the HMD shows, but is slightly higher, likely a result of selection via socioeconomic status, a limitation that will be discussed later in this article. Ideally, we would have used younger age cutoffs to show these trends, but due to the United States not having as long-running demographic records on life expectancy and mortality (i.e., Sweden), we opted to use older age cutoffs to add validity to our estimates before moving forward. Moreover, to show transparency in our estimates, and highlight the fact that we are using what is undoubtedly a highly selected sample, we also compared our estimates in 1900 to those of the United States Social Security Administration, which showed much higher biases at younger ages (Table SB, Appendix, Supplementary material). We also compared our estimates to those of Hacker (2010) for the white population of the century, which also revealed much of the same bias (Table SC, Appendix, Supplementary material).

Table 1. Descriptive statistics of sample by birth cohort

Birth Decade	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890
<i>Males</i>										
Longevity (SD)	68.5 (16.8)	68.0 (16.9)	67.4 (17.8)	67.0 (18.7)	67.4 (18.8)	68.5 (17.8)	68.8 (17.6)	69.0 (18.0)	69.2 (18.0)	69.5 (17.9)
Lifespan variation (SD)	18.8 (2.8)	19.0 (2.5)	20.3 (2.5)	21.5 (2.4)	22.0 (2.7)	22.2 (2.4)	22.4 (2.8)	22.7 (2.5)	22.6 (2.1)	22.3 (1.8)
# of Children (SD)	3.2 (1.8)	3.3 (1.8)	3.3 (1.8)	3.3 (1.7)	3.37 (1.7)	3.4 (1.7)	3.3 (1.7)	3.2 (1.7)	3.2 (1.6)	3.0 (1.6)
N = 633,858	35,047	39,099	42,182	45,040	51,857	61,483	67,793	81,502	96,964	112,887
<i>Females</i>										
Longevity (SD)	66.9 (19.1)	66.1 (19.5)	66.2 (19.8)	66.3 (20.0)	66.5 (20.1)	67.1 (20.0)	68.3 (19.9)	69.7 (20.2)	71.8 (20.1)	74.6 (19.7)
Lifespan variation (SD)	20.9 (2.9)	21.3 (2.6)	21.9 (2.6)	22.7 (2.6)	23.3 (2.7)	23.7 (2.1)	23.9 (2.4)	24.6 (2.2)	24.3 (1.9)	24.1 (1.7)
# of Children (SD)	3.2 (1.8)	3.2 (1.8)	3.3 (1.8)	3.3 (1.7)	3.4 (1.7)	3.4 (1.7)	3.4 (1.7)	3.3 (1.7)	3.2 (1.6)	3.1 (1.6)
N = 527,611	25,604	28,601	32,047	35,854	41,789	51,692	59,352	71,124	84,324	97,220
<i>Full sample</i>										
Longevity (SD)	67.8 (17.9)	67.2 (18.0)	66.9 (18.7)	66.7 (19.3)	67.0 (19.4)	67.9 (18.8)	68.6 (18.7)	69.4 (19.0)	70.4 (19.0)	71.9 (19.0)

(Continued)

Table 1. (*Continued*)

Birth Decade	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890
Lifespan variation	19.7	19.9	21.0	22.0	22.6	22.9	23.1	23.6	23.4	23.1
(SD)	(3.0)	(2.8)	(2.7)	(2.5)	(2.8)	(2.4)	(2.7)	(2.5)	(2.1)	(2.0)
# of Children (SD)	3.2	3.3	3.3	3.3	3.4	3.4	3.3	3.3	3.2	3.0
	(1.8)	(1.8)	(1.8)	(1.7)	(1.7)	(1.7)	(1.7)	(1.7)	(1.6)	(1.6)
N = 1,161,469	60,658	67,701	74,229	80,894	93,946	113,175	127,145	152,626	181,288	210,107

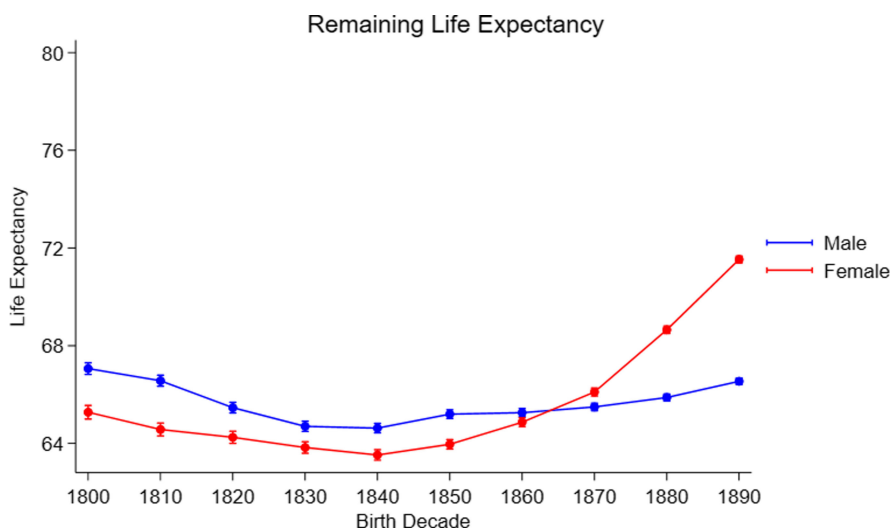


Figure 2. Predicted life expectancy conditional on survival to age 5.

Table 1 displays the descriptive statistics of our sample. Among the 1,161,469 individuals in our sample, 633,858 (54.6 percent) were males and 527,611 (45.4 percent) were females. Figure A5 (see Appendix, Supplementary material) shows the gender distribution over time, with males making up more of the sample in earlier cohorts. Due to conditioning on survival to age 5, our longevity estimates ranged from 66.7 to 71.9 across the decades. In terms of number of children, the average number of children an individual had in our sample ranged from 3.0 children in 1890 to 3.4 among mid-century birth cohorts, and the average lifespan variation ranged from 19.7 to 23.6. For sex-specific descriptive results, we next delve more into the trends across birth cohorts by gender in the subsequent figures.

For our cohort life expectancy estimates, Figure 1A (see Appendix, Supplementary material) shows the general trends for males and females that survive to ages 5, 25, 50, and 75 years of age. For age 5, it shows three regimes of change that take place throughout the time period. From 1800 to 1830, the trend is relatively flat for males and females. Then, from 1830 to about 1860/1870, there is some acceleration for females, in which their life expectancy start to converge with males. After this, from 1870 onward, there is a rapid acceleration. This trend is largely the same for males and females who survive to age 25, albeit a narrower gap existed from 1800 to about 1860. As for conditioning on survival to the higher ages of 50 and 70, there is a consistent female advantage throughout the entire period, while also showcasing an acceleration in the gap between males and females from 1860 onwards.

Figures 2 and 3 show results from models that regress cohort longevity on birth decades from 1800 to 1890, with 95% confidence intervals. The former of the figures looks at our baseline estimates and the latter controls for the state of birth. Both figures show relatively similar trends to our previous results, with males having a slight advantage until mid-century, only to lose it to women, who see a sharp

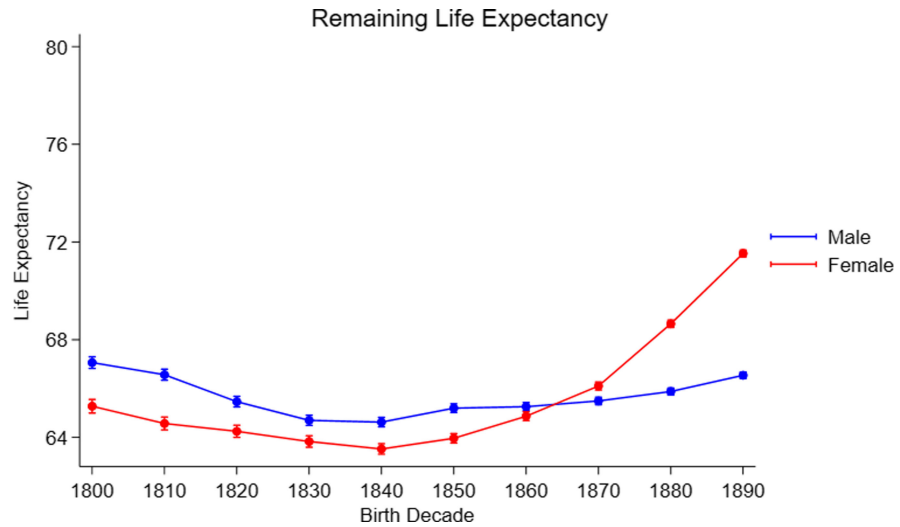


Figure 3. Predicted life expectancy conditional on survival to age 5 (controlling for state of birth).

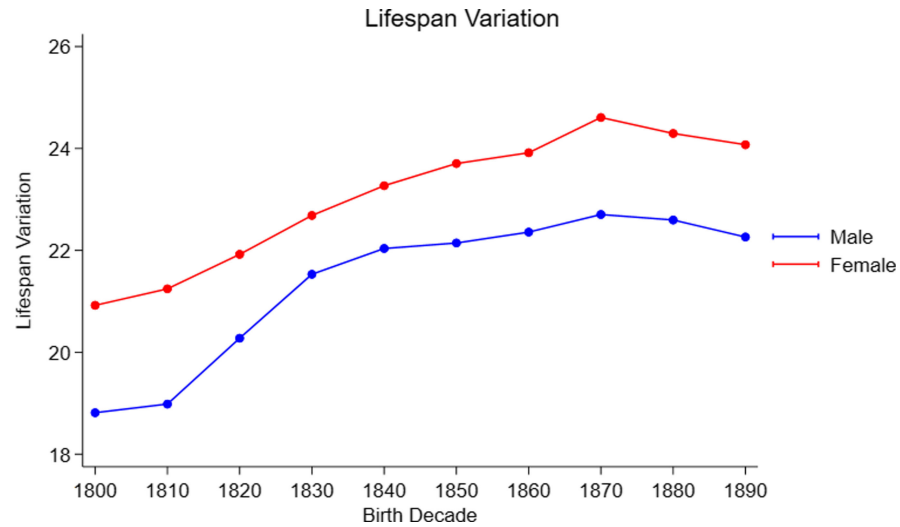


Figure 4. Lifespan variation conditional on survival to age 5.

acceleration in longevity in later years. Interestingly, there is little change in estimates when the control for the state of birth is incorporated, despite some states showcasing an association with longevity. As a robustness check, we again estimated the prior models, but this time we look at longevity conditional on survival to age 10, which show similar estimates (see Figures A3 and A4, respectively, Appendix, Supplementary material).

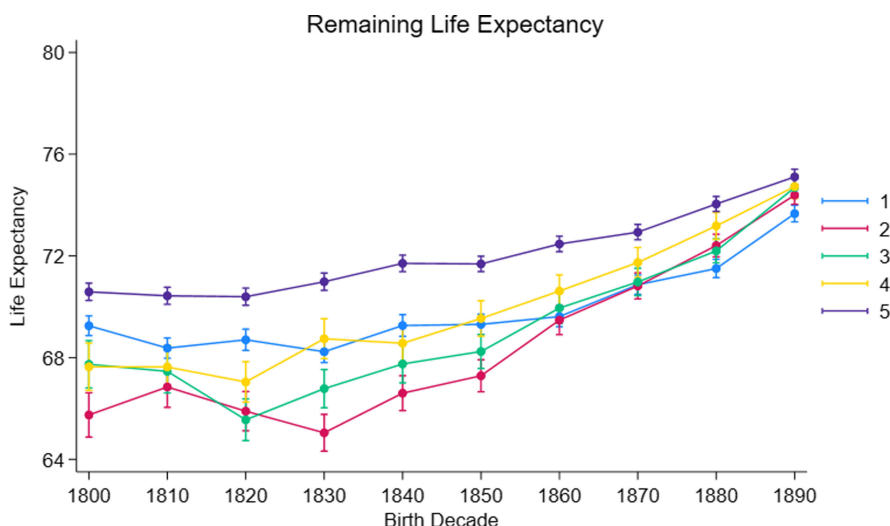


Figure 5. Predicted life expectancy at age 15 by number of children.

Moving on to trends in the dispersion of longevity, we display lifespan variation estimates in Figure 4. From 1800 to 1840, both males and females see a rise in variation, with males having a steeper increase. After this period, there is relative stagnation in the amount of lifespan variability until 1870, after which a decline begins, until both male and female lifespan variation converges at the end of the century. One noteworthy finding from these results is that throughout the entire century, our estimates show that females had higher lifespan variation than males, which contradicts previous literature (Aburto et al., 2020; Colchero et al., 2016). However, it should be noted that previous literature has not considered the context of the United States in the nineteenth century and thus, our work offers a novel contribution to this literature. To test the robustness of this finding, we also estimate lifespan variation conditional on survival to age 50 (see Figure A4 and Table A3, Appendix, Supplementary material). Ultimately, we found consistent but narrowed gaps.

With regard to the role that childbearing has on longevity, Figure 5 shows the predicted longevity that individuals who survive to 15 have by number of children. Broadly, it shows in our smaller sample of those who have children that those who have five or more see the highest longevity. However, one issue with interpreting this finding is that females are positively selected; for instance, they would have needed to survive the first three children in order to have a fourth child. Nevertheless, one striking finding of the trends observed here is that in 1800, there was a great deal of variability in the longevity estimates, which began to collapse around 1860, and then accelerated. To capture the window of childbearing, in Figure 6 we display the same estimates but limit it to those that survive to age 50. We note that there is far less variability once that is accounted for, with the variability being limited to the earlier half of the century.

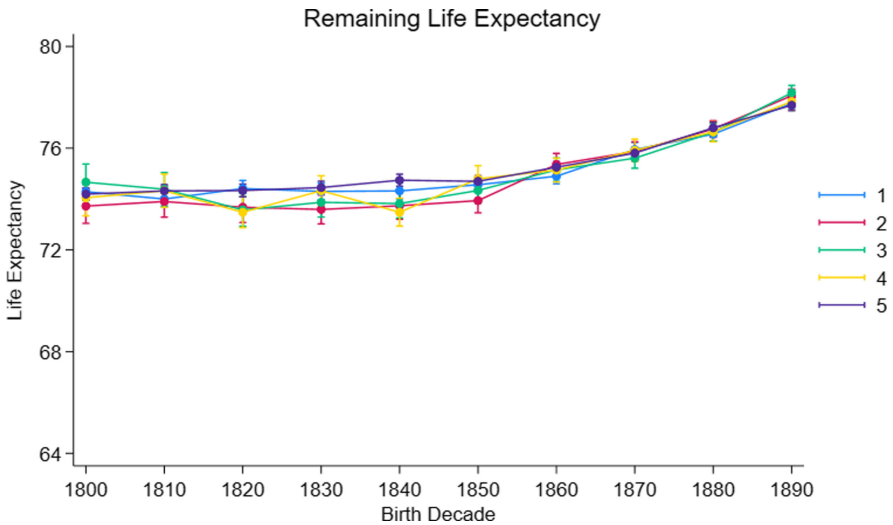


Figure 6. Predicted life expectancy at age 50 by number of children.

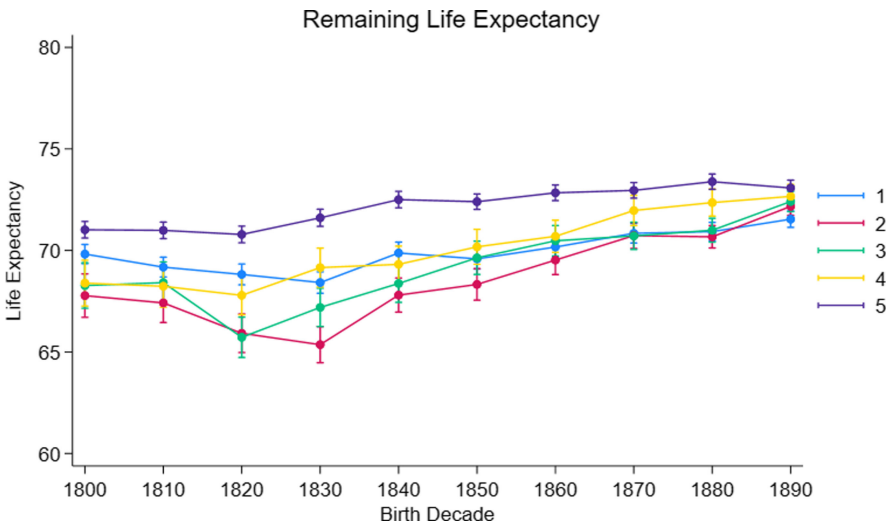


Figure 7. Predicted life expectancy at age 15 by number of children, males.

Figures 7 and 8 show the trends presented in Figure 5 but look at male and female longevity, respectively. At first glance, it appears that the first half of the century saw greater variation in how long one lived by the number of children they had. Then, by the mid-century, this variation largely collapsed, especially for the female population. While males see relatively stable trends, females see a steep increase in their longevity, up to 15 years for some estimates, suggesting major improvements to longevity. Conditioning on surviving to 50 years of age

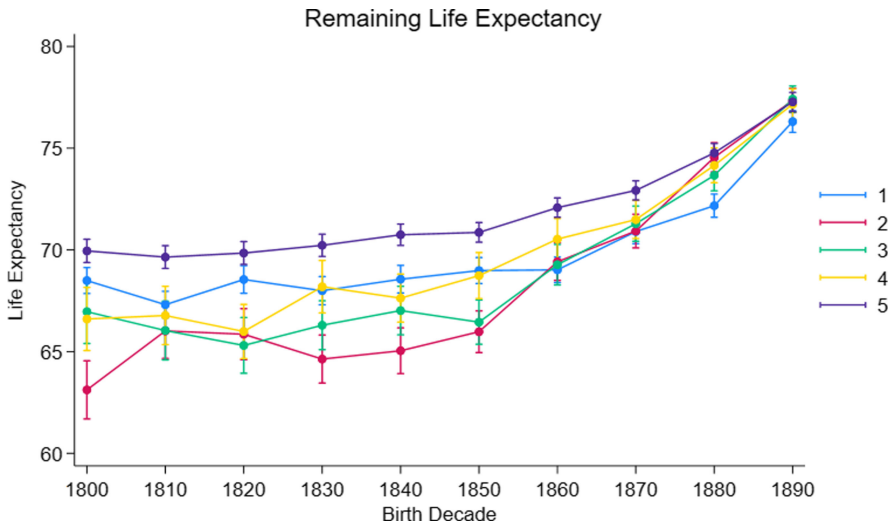


Figure 8. Predicted life expectancy at age 15 by number of children, females.

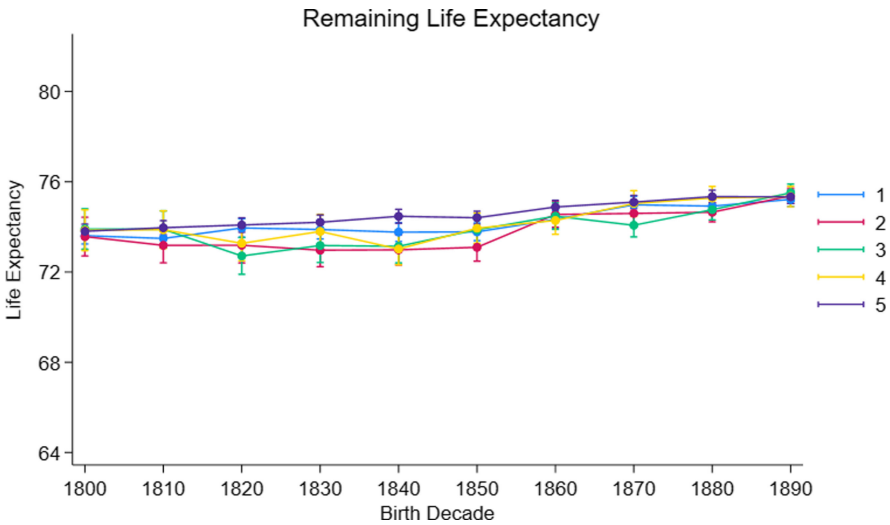


Figure 9. Predicted life expectancy at age 50 by number of children, males.

(Figures 9 and 10) does relatively little for males with regard to longevity, whereas it makes a large difference for females (see Table A4, Appendix, Supplementary material). However, we should note that in the case of females, this finding is not entirely unexpected, given that future fertility is dependent on prior fertility. In other words, a woman who has more than five children will, in theory, need to live longer to have those five children compared to a woman who only gives birth to one child. Yet, despite this limitation, it is important to note that even those who have

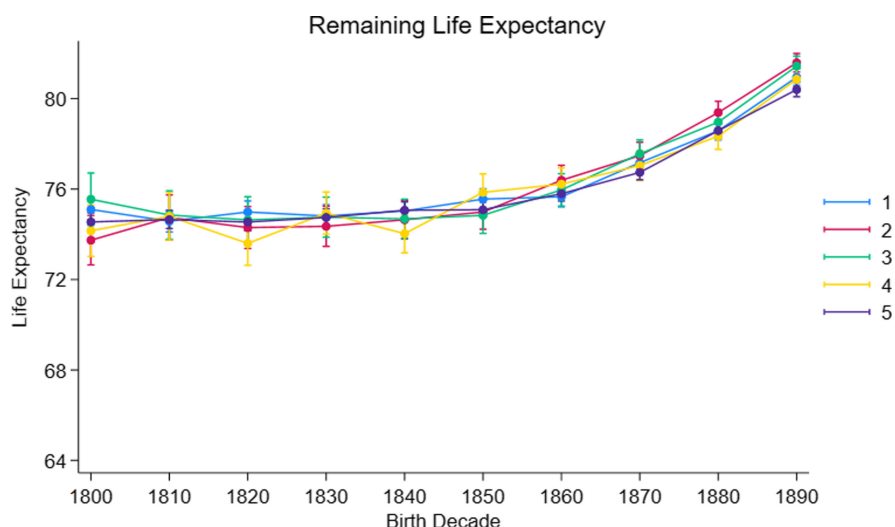


Figure 10. Predicted life expectancy at age 50 by number of children, females.

fewer children see similar longevity estimates to those who have more. Given these limitations, we also opted to look at fertility within couples in order to factor out much of the unobserved characteristics that may be driving findings, which did not yield dramatically different results (See Figures A7 and A8, Appendix, Supplementary material). These findings indirectly support a mechanism regarding fertility that plays a role in longevity differences between the sexes. One potential mechanism could be improvements to infant mortality, which would largely allow females to forego future pregnancies that in turn would eliminate the potential for maternal mortality during the window of childbearing, given that more infants survive to childhood and potentially adulthood.

To summarize these figures, the results suggest that the decrease in longevity variation is largely due to the increase in longevity for those with fewer children, which may be due to two reasons: (i) the decrease in death selection (i.e., the decrease in early deaths lead to a weaker association between longevity and the number of children); (ii) the decrease in the disadvantage of having fewer children. Following up on these possible explanations, the findings suggest a large reduction in longevity variation when conditioning on survival to 50, implying that large longevity variation in the first half of the 19th century is mainly due to a higher amount of deaths before 50. We also see that, even when conditioning on survival to 50, the increase in longevity among those with fewer children (especially women with two children) suggests evidence for the second mechanism. Finally, even when conditioning on survival to 50 and stratified by the number of children, we see a steep and consistent increase in female longevity in the second half of the nineteenth century, implying that the trends remain largely unexplained.

To describe the patterns of the increase in female longevity in more detail, we performed a decomposition analysis (see Table A1, Appendix, Supplementary material). First, results show that the endowment effects are negative, which became

smaller and negligible when conditioning on survival to 50, especially among females. The results imply that, assuming that the relationship between longevity and fertility is fixed at <1870, the decrease in fertility itself has little impact on the increase in female longevity. Second, the coefficient effects are close to zero and not statistically significant, especially when conditioning on survival to 50, meaning that the changing relationship between fertility and longevity itself does not account for the increase in female longevity. Third, the interaction effects – the joint effects of the changes in the level and coefficients of fertility – are statistically significant among all groups, including those who survived 50 years and especially among females. This implies that the joint changes in the distribution and effects of fertility account for about 5.5 percent of the observed increase in female longevity among those who survive to 50. In sum, the decomposition reveals that the decrease in fertility and increase in longevity among those with fewer children partly explain the increase in longevity, further adding evidence of the role fertility played in sex longevity differences in this historical period.

Discussion

Prior work has thoroughly documented sex differences in survival. Specifically, prior work has documented that historically females have outlived males in almost every country in the world throughout history (Austad, 2006), whereas others specifically point to examples where males outlived females in countries such as the Netherlands and Italy in different periods in the 1800s (Barford et al., 2006). Yet, there are relatively few contributions to this literature in the context of the United States, and less specifically during the period of the nineteenth century, during which there was some evidence that a slight male advantage existed for specific age groups or ranges (Goldin and Lleras-Muney, 2019). Limitations associated with such an undertaking include a lack of adequate or reliable data on mortality, which the United States did not begin to keep until well after the century concluded (Hacker, 2010). Other issues relate to the lack of contextual information, such as place of birth or fertility information.

This study aimed to explore how sex differences in longevity changed in the United States throughout the nineteenth century. In looking at birth cohorts across the century, we found that until the 1860–1870s, males held an advantage over females in terms of longevity. Yet, after this mid-century period, subsequent birth cohorts saw the emergence of a female advantage in longevity, one that has persisted through to the present day. As touched upon in the results section – one potential mechanism for this rather sharp increase in female longevity could be reductions in maternal mortality. With greater proportions of children surviving to older ages, it could potentially remove the desire to have more children as a way to account for potential premature mortality. This, in turn, would cumulatively reduce the instances in which a female would be pregnant, thereby reducing opportunities for maternal mortality to occur. Prior research looking at the United Kingdom found that maternal mortality was quite high throughout this century, and that by the century's end, the rate was half of what the United States rate was, providing further evidence as a mechanism driving these findings (Chamberlain, 2006; Goldenberg

and McClure, 2011). Another potential mechanism could be specifically the public health reforms that the United States underwent starting in the 1850s. Indeed, many health reformers in the latter half of the 1800s specifically sought to tackle the idea that ill health and by extension mortality were a result of the environment and hygiene rather than economic conditions, and they pushed for better sanitation standards (Meckel, 2015).

Using unique and massive individual data from our genealogic data source, the results in this article make many novel contributions to the literature while supporting previous research. First, this study supports prior scholars in affirming that males held a slight advantage over females in longevity up until mid-century, which in the United States, coincided with the Civil War, an event that directly had an impact not only on male mortality but also many indirect effects on female mortality. However, such studies are limited in the sense that they either have data on only a single state (Goldin and Lleras-Muney, 2019) or a racial group (Hacker, 2010). Yet, another plausible explanation for this result could also be that the observed effect also coincided with the increased number of smoking-attributable deaths among men, as found in prior work (Wensink *et al.*, 2020). Indeed, prior work has shown that despite not being as widely used by individuals during this time period, the impact of smoking was higher among birth cohorts in the late 1800s compared to birth cohorts now (Oza *et al.*, 2011). This reinforces the notion that excess mortality among males during that time allowed females to gain an initial advantage. Second, this article revealed some evidence of spatial variations in these patterns, in the form of state of birth. However, the introduction of state of birth in the analysis did not attenuate the estimates of birth decade on longevity a great deal, as evidenced in our figures.

Third, our results revealed that females had higher lifespan variation during this period than males, which is novel in the sense that it has been frequently found that males have higher lifespan variation (Aburto *et al.*, 2020; Colchero *et al.*, 2016). It should be noted that prior studies that examine lifespan variation between males and females primarily look at contexts outside of the United States or focus on the United States from the 20th century onward. This finding of higher lifespan variation among females speaks to the underlying disparities between males and females during this period, with females showing higher dispersion even after they gained and grew an advantage over males in longevity. Moreover, since this has not been as thoroughly examined in the context of the pre-20th century United States, this could represent a snapshot of trends present before administrative data on mortality were available. We further examined this phenomenon by looking at the distributions of each birth cohort. Ultimately, we found that the increased proportion of deaths taking place prior to 50 years old in the female population in earlier cohorts is what was likely driving this (see Appendix, Figure A6).

Fourth, this study reveals the role fertility played in longevity during the 19th century, and how survival to later years, an age at which the window of childbearing closes for many, makes a noteworthy impact. We illustrate that for males and females, there is great variation in longevity depending on the number of children they have, which is substantially reduced in the latter half of the century. However, once survival to age 50 is accounted for, there is a collapse of that variation, for both males and females, but greater for the latter, and a rather sharp rise in longevity

across the century for successive cohorts. Essentially, our findings indirectly support the influence of fertility on longevity and explain overall trends in longevity. This contribution is critical, given that previous work has looked at the longevity of those in this period only in ages above 40 (Beltrán-Sánchez et al., 2015), essentially missing this story.

That said, there are important limitations in this study that we acknowledge. First, while the genealogical data used in this study cover long historical periods and provide a promising avenue for demographers to use (Alburez-Gutierrez et al., 2019), the longevity data come from a selective sample, and thus may have longer lifespans than the general population at that time. This can be seen in the fact that in the overall sample, there are more males than females (54.6 percent vs 45.4 percent, respectively). During this time period, those who were more socioeconomically advantaged had better records and were disproportionately likely to be male, in part due to the heritability of wealth, which could explain some of these findings (Stelter and Alburez-Gutierrez, 2022).

Another limitation of this work is that Geni data do not provide information on those whose child information is missing, thus limiting available data on the number of children to about slightly less than half of our sample. Furthermore, those who are childless are also marked as “missing” information in the data. Thus, our estimates may underestimate the effects of fertility on longevity due to not having full information on childbearing for individuals in the sample. For instance, it could be the case that individuals who did not have any children lived longer than those who had any, for said individuals likely would not have had to expend any resources. We also did not possess rich demographic information on individuals in order to incorporate a sample selection correction to account for this, and as such, opted to focus solely on fertility from the perspective of those who did have children. Future research should aim to consider the fertility of populations during this time to see if childlessness improved longevity or if the conclusions we reached with a number of children are robust. Related to this, we did not possess data on marriages and remarriages. This data limitation prevents us from exploring the role of marriage markets, the shared mortality risks, and the disrupted fertility from lack of a partner as a result of widowhood, all of which may represent a type of mediation that may be driving the associations we see in this article. As such, future studies would benefit from leveraging data on marriage and remarriage in order to see the role it plays in these processes, if any.

A third limitation of our study stems from a lack of data on ethnic background. Given the racial makeup of the United States during this period, along with historical processes playing out that disproportionately revolved around race, future contributions to this literature would consider how racial and ethnic background played a role in longevity estimates. This examination extends to those of different ethnic backgrounds as well, for certain backgrounds were more likely to be considered “white” than others. However, we neglected that information by controlling for only those born in the United States. Despite these limitations, we believe that this study is unique due to the fact that our rich dataset enables us both to calculate individual lifespans, which allows us to not be as reliant on period measures while also being able to factor in controls for geography and fertility.

A fourth limitation relates to the overall generalizability of our analytic sample. As stated above, the fact that the individuals in these data are likely a very select sample, along with the fact that a large portion of the 1.2 million respondents are removed when looking at those with children, it is important to note that our sample is not fully representative of the United States during this time period. Given these validity concerns initially, we compared our data with those from the HMD, both with the full analytic sample, and those with child information, and our estimates yielded similar findings. That being said, we would again emphasize caution in interpreting these findings. Despite these data not being wholly representative of the United States population at this time, we would still highlight that the data used in this article are unique in the sense that they complement and help to improve on other sources of data (HMD; Hacker, 2010; Goldin and Lleras-Muney, 2019) used by historical demographers and social scientists. Future research should work to incorporate the data used in this study with others to present valid findings on demographic processes from birth cohorts during this time period and others.

Conclusion

The rise in life expectancy over the centuries signifies a remarkable feat in human history (Oeppen and Vaupel, 2002). This study is one of the first to use online genealogical data to study a fundamental, yet recent, demographic phenomenon: the female advantage in life expectancy. Specifically, this article is noteworthy given that it observes this phenomenon in the context of the United States, whereas prior work has exclusively been focused on other settings, mainly in Western and Northern Europe. Ultimately, we find that prior to the mid-nineteenth century, males held a slight advantage over females, after which females gained and maintained this advantage. In addition to this, we also found that individuals who had more children saw benefits to their longevity. Yet, by the end of the century, there were no substantial differences in the number of children an individual had. Moreover, we discovered that even though the female advantage in longevity emerged mid-century, females had greater lifespan variation throughout the entire time period, in part due to a greater percentage of deaths taking place at maternal ages. In short, this research helps to paint a picture of trends in longevity throughout a time period for which we need to rely on more non-traditional methods to understand population dynamics.

Acknowledgements. This research was carried out using the facilities of the Center for Demography and Ecology and Center, supported by Eunice Kennedy Shriver National Institute of Child Health and Human Development grant P2C HD047873 and the Center for Demography of Health and Aging, which is supported by National Institute on Aging grant P30 AG017266. Topping was funded in part by the National Institute on Aging training grant T32 AG00129. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank the participants of the 2022 Southern Demographic Association Annual Meeting for helpful comments on this paper.

References

- Aburto, José Manuel, Francisco Villavicencio, Ugofilippo Basellini, Søren Kjærgaard, and James W. Vaupel (2020) "Dynamics of life expectancy and life span equality," *Proceedings of the National Academy of Sciences* 117 (10): 5250–59. <https://doi.org/10.1073/pnas.1915884117>
- Alburez-Gutierrez, Diego, Emilio Zagheni, Samin Aref, Sofia Gil-Clavel, André Grow, and Daniela Veronica Negraia (2019) "Demography in the Digital Era: New Data Sources for Population Research," Preprint. SocArXiv. <https://doi.org/10.31235/osf.io/24jp7>
- Austad, Steven N. (2006) "Why women live longer than men: Sex differences in longevity," *Gender Medicine* 3 (2): 79–92. [https://doi.org/10.1016/S1550-8579\(06\)80198-1](https://doi.org/10.1016/S1550-8579(06)80198-1)
- Austad, Steven N., and Kathleen E. Fischer (2016) "Sex differences in lifespan," *Cell Metabolism* 23 (6): 1022–33. <https://doi.org/10.1016/j.cmet.2016.05.019>
- Barford, Anna, Danny Dorling, George Davey Smith, and Mary Shaw (2006) "Life expectancy: women now on top everywhere," *BMJ* 332 (7545): 808. <https://doi.org/10.1136/bmj.332.7545.808>
- Beltrán-Sánchez, Hiram, Caleb E. Finch, and Eileen M. Crimmins (2015) "Twentieth century surge of excess adult male mortality," *Proceedings of the National Academy of Sciences* 112 (29): 8993–98. <https://doi.org/10.1073/pnas.1421942112>
- Bergeron-Boucher, Marie-Pier, Jesús-Adrian Alvarez, Ilya Kashnitsky, Virginia Zarulli, and James W. Vaupel (2020) "Not all females outlive all males: A new perspective on lifespan inequalities between sexes," Preprint. SocArXiv. <https://doi.org/10.31235/osf.io/typws>
- Bllell, Mwenza (2017) "Grandmother Hypothesis, Grandmother Effect, and Residence Patterns," in Hilary Callan (ed.) *The International Encyclopedia of Anthropology*, 1st ed. Wiley: 1–5. <https://doi.org/10.1002/9781118924396.wbiea2162>
- Boing, Antonio Fernando, Alexandra Crispim Boing, Jack Cordes, Rockli Kim, and S. V. Subramanian (2020) "Quantifying and explaining variation in life expectancy at census tract, county, and state levels in the United States," *Proceedings of the National Academy of Sciences* 117 (30): 17688–94. <https://doi.org/10.1073/pnas.2003719117>
- Bronnum-Hansen, H., and K. Juel (2001) "Abstinence from smoking extends life and compresses morbidity: a population based study of health expectancy among smokers and never smokers in Denmark," *Tobacco Control* 10 (3): 273–78. <https://doi.org/10.1136/tc.10.3.273>
- Chamberlain, Geoffrey (2006) "British Maternal Mortality in the 19th and Early 20th Centuries," *Journal of the Royal Society of Medicine* 99 (11): 559–63. <https://doi.org/10.1177/014107680609901113>
- Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler (2016) "The Association Between Income and Life Expectancy in the United States, 2001–2014," *JAMA* 315 (16): 1750. <https://doi.org/10.1001/jama.2016.4226>
- Colchero, Fernando, Roland Rau, Owen R. Jones, Julia A. Barthold, Dalia A. Conde, Adam Lenart, Laszlo Nemeth, et al. (2016) "The emergence of longevous populations," *Proceedings of the National Academy of Sciences* 113 (48): E7681–90. <https://doi.org/10.1073/pnas.1612191113>
- Condran, Gretchen A., and Eileen Crimmins-Gardner (1978) "Public health measures and mortality in U.S. cities in the late nineteenth century," *Human Ecology* 6 (1): 27–54. <https://doi.org/10.1007/BF00888565>
- Crimmins, Eileen M., Hyunju Shim, Yuan S. Zhang, and Jung Ki Kim (2019) "Differences between men and women in mortality and the health dimensions of the morbidity process," *Clinical Chemistry* 65 (1): 135–45. <https://doi.org/10.1373/clinchem.2018.288332>
- Daymont, Thomas N., and Paul J. Andrisani (1984) "Job preferences, college major, and the gender gap in earnings," *The Journal of Human Resources* 19 (3): 408. <https://doi.org/10.2307/145880>
- Deryugina, Tatyana, and David Molitor (2021) "The causal effects of place on health and longevity," *Journal of Economic Perspectives* 35 (4): 147–70. <https://doi.org/10.1257/jep.35.4.147>
- Drevenstedt, Greg L., Eileen M. Crimmins, Sarinnapha Vasunilashorn, and Caleb E. Finch (2008) "The rise and fall of excess male infant mortality," *Proceedings of the National Academy of Sciences* 105 (13): 5016–21. <https://doi.org/10.1073/pnas.0800221105>

- Edwards, Ryan D., and Shripad Tuljapurkar (2005) "Inequality in life spans and a new perspective on mortality convergence across industrialized countries," *Population and Development Review* 31 (4): 645–74. <https://doi.org/10.1111/j.1728-4457.2005.00092.x>
- Ehrlich, Shelley (2015) "Effect of fertility and infertility on longevity," *Fertility and Sterility* 103 (5): 1129–35. <https://doi.org/10.1016/j.fertnstert.2015.03.021>
- Goldenberg, Robert L., and Elizabeth M. McClure (2011) "Maternal mortality," *American Journal of Obstetrics and Gynecology* 205 (4): 293–95. <https://doi.org/10.1016/j.ajog.2011.07.045>
- Goldin, Claudia, and Adriana Lleras-Muney (2019) "XX > XY?: The changing female advantage in life expectancy," *Journal of Health Economics* 67 (September): 102224. <https://doi.org/10.1016/j.jhealeco.2019.102224>
- Hacker, J. David (2010) "Decennial life tables for the white population of the United States, 1790–1900," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 43 (2): 45–79. <https://doi.org/10.1080/01615441003720449>
- Haines, Michael (2001) *The Urban Mortality Transition in the United States, 1800–1940*. Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/h0134>
- Hayward, Mark D., Amy M. Pienta, and Diane K. McLaughlin (1997) "Inequality in men's mortality: The socioeconomic status gradient and geographic context," *Journal of Health and Social Behavior* 38 (4): 313. <https://doi.org/10.2307/2955428>
- Herd, P., D. Carr, and C. Roan (2014) "Cohort profile: Wisconsin longitudinal study (WLS)," *International Journal of Epidemiology* 43 (1): 34–41. <https://doi.org/10.1093/ije/dys194>
- Janssen, Fanny (2020) "Changing contribution of smoking to the sex differences in life expectancy in Europe, 1950–2014," *European Journal of Epidemiology* 35 (9): 835–41. <https://doi.org/10.1007/s10654-020-00602-x>
- Kalben, Barbara Blatt (2000) "Why men die younger: Causes of mortality differences by sex," *North American Actuarial Journal* 4 (4): 83–111. <https://doi.org/10.1080/10920277.2000.10595939>
- Kaplanis, Joanna, Assaf Gordon, Tal Shor, Omer Weissbrod, Dan Geiger, Mary Wahl, Michael Gershovits, et al. (2018) "Quantitative analysis of population-scale family trees with millions of relatives," *Science* 360 (6385): 171–75. <https://doi.org/10.1126/science.aam9309>
- Lindahl-Jacobsen, Rune, Heidi A. Hanson, Anna Oksuzyan, Geraldine P. Mineau, Kaare Christensen, and Ken R. Smith (2013) "The male-female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910," *Annals of Epidemiology* 23 (4): 161–66. <https://doi.org/10.1016/j.annepidem.2013.02.001>
- Luy, Marc, and Katrin Gast (2014) "Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy," *Gerontology* 60 (2): 143–53. <https://doi.org/10.1159/000355310>
- Lycett, J. E., R. I. Dunbar, and E. Voland (2000) "Longevity and the costs of reproduction in a historical human population," *Proceedings of Biological Sciences* 267 (1438): 31–35. <https://doi.org/10.1098/rspb.2000.0962>
- Mann, David R., and Todd Honeycutt (2016) "Understanding the disability dynamics of youth: health condition and limitation changes for youth and their influence on longitudinal survey attrition," *Demography* 53 (3): 749–76. <https://doi.org/10.1007/s13524-016-0469-7>
- McArdle, Patrick F., Toni I. Pollin, Jeffrey R. O'Connell, John D. Sorkin, Richa Agarwala, Alejandro A. Schäffer, Elizabeth A. Streeten, Terri M. King, Alan R. Shuldiner, and Braxton D. Mitchell (2006) "Does having children extend life span? A genealogical study of parity and longevity in the Amish," *The Journals of Gerontology: Series A* 61 (2): 190–95. <https://doi.org/10.1093/gerona/61.2.190>
- Meckel, Richard A. (2015) *Save the Babies: American Public Health Reform and the Prevention of Infant Mortality, 1850–1929*. Rochester, NY: Univ. of Rochester Press.
- Montez, Jennifer Karas, Jason Beckfield, Julene Kemp Cooney, Jacob M. Grumbach, Mark D. Hayward, Huseyin Zeyd Koytak, Steven H. Woolf, and Anna Zajacova (2020) "US State Policies, Politics, and Life Expectancy," *The Milbank Quarterly* 98 (3): 668–99. <https://doi.org/10.1111/1468-0009.12469>
- Mosca, Lori, Elizabeth Barrett-Connor, and Nanette Kass Wenger (2011) "Sex/gender differences in cardiovascular disease prevention: What a difference a decade makes," *Circulation* 124 (19): 2145–54. <https://doi.org/10.1161/CIRCULATIONAHA.110.968792>

- Muller, H.-G., J.-M. Chiou, J. R. Carey, and J.-L. Wang (2002) "Fertility and life span: Late children enhance female longevity," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 57 (5): B202–6. <https://doi.org/10.1093/gerona/57.5.B202>
- Oeppen, Jim, and James W. Vaupel (2002) "Broken limits to life expectancy," *Science* 296 (5570): 1029–31. <https://doi.org/10.1126/science.1069675>
- Oksuzyan, Anna, Knud Juel, James W. Vaupel, and Kaare Christensen (2008) "Men: good health and high mortality. Sex differences in health and aging," *Aging Clinical and Experimental Research* 20 (2): 91–102. <https://doi.org/10.1007/BF03324754>
- Oza, Shefali, Michael J. Thun, S. Jane Henley, Alan D. Lopez, and Majid Ezzati (2011) "How many deaths are attributable to smoking in the United States? Comparison of methods for estimating smoking-attributable mortality when smoking prevalence changes," *Preventive Medicine* 52 (6): 428–33. <https://doi.org/10.1016/j.ypmed.2011.04.007>
- Pope, Clayne L. (1992) "Adult Mortality in America before 1900: A View from Family Histories" *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*. Chicago: University of Chicago Press: 267–96.
- Preston, Samuel H., and Michael R. Haines (1991) *Fatal Years: Child Mortality in Late Nineteenth-Century America*. Princeton, NJ: Princeton University Press.
- Preston, Samuel H., and Haidong Wang (2006) "Sex mortality differences in The United States: The role of cohort smoking patterns," *Demography* 43 (4): 631–46. <https://doi.org/10.1353/dem.2006.0037>
- Rigby, J. E, and D. Dorling (2007) "Mortality in relation to sex in the affluent world," *Journal of Epidemiology & Community Health* 61 (2): 159–64. <https://doi.org/10.1136/jech.2006.047381>
- Rogers, Richard G., Bethany G. Everett, Jarron M. Saint Onge, and Patrick M. Krueger (2010) "Social, behavioral, and biological factors, and sex differences in mortality," *Demography* 47 (3): 555–78. <https://doi.org/10.1353/dem.0.0119>
- Sasson, Isaac (2016) "Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010," *Demography* 53 (2): 269–93. <https://doi.org/10.1007/s13524-015-0453-7>
- Schoeni, Robert F., and Emily E. Wiemers (2015) "The implications of selective attrition for estimates of intergenerational elasticity of family income," *The Journal of Economic Inequality* 13 (3): 351–72. <https://doi.org/10.1007/s10888-015-9297-z>
- Seifarth, Joshua E., Cheri L. McGowan, and Kevin J. Milne (2012) "Sex and life expectancy," *Gender Medicine* 9 (6): 390–401. <https://doi.org/10.1016/j.genm.2012.10.001>
- Stelter, Robert, and Diego Albrez-Gutierrez (2022) "Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics," *Proceedings of the National Academy of Sciences* 119 (10): e2120455119. <https://doi.org/10.1073/pnas.2120455119>
- Sundberg, Louise, Neda Agahi, Johan Fritzell, and Stefan Fors (2018) "Why is the gender gap in life expectancy decreasing? The impact of age- and cause-specific mortality in Sweden 1997–2014," *International Journal of Public Health* 63 (6): 673–81. <https://doi.org/10.1007/s00038-018-1097-3>
- Weidner, Gerdi (2000) "Why do men get more heart disease than women? An international perspective," *Journal of American College Health* 48 (6): 291–94. <https://doi.org/10.1080/07448480009596270>
- Wensink, Maarten, Jesús-Adrián Alvarez, Silvia Rizzi, Fanny Janssen, and Rune Lindahl-Jacobsen (2020) "Progression of the smoking epidemic in high-income regions and its effects on male-female survival differences: a cohort-by-age analysis of 17 countries," *BMC Public Health* 20 (1): 39. <https://doi.org/10.1186/s12889-020-8148-4>
- Zafeiris, Konstantinos N. (2020) "Gender differences in life expectancy at birth in Greece 1994–2017," *Journal of Population Research* 37 (1): 73–89. <https://doi.org/10.1007/s12546-019-09239-4>
- Zarulli, Virginia, Julia A. Barthold Jones, Anna Oksuzyan, Rune Lindahl-Jacobsen, Kaare Christensen, and James W. Vaupel (2018) "Women live longer than men even during severe famines and epidemics," *Proceedings of the National Academy of Sciences* 115 (4): E832–40. <https://doi.org/10.1073/pnas.1701535115>