I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16018

Trends in the Female Longevity Advantage of 19th-Century Birth Cohorts: Exploring the Role of Place and Fertility

Jason Fletcher
Michael Topping
Won-tak Joo

MARCH 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16018

Trends in the Female Longevity Advantage of 19th-Century Birth Cohorts: Exploring the Role of Place and Fertility

Jason Fletcher
University of Wisconsin-Madison and IZA

Won-tak Joo
University of Wisconsin-Madison

Michael Topping
University of Wisconsin-Madison

MARCH 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Trends in the Female Longevity Advantage of 19th-Century Birth Cohorts: Exploring the Role of Place and Fertility*

This paper uses massive online genealogy data from the United States over the 19th century to estimate period and cohort-based sex differences in longevity. Following previous work, we find a longevity reversal in the mid-19th 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 allows analysis of spatial patterns and of 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.

JEL Classification: J11

Keywords: longevity, sex differences, US, genealogy

Corresponding author:
Jason Fletcher
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706
USA
E-mail: jason.fletcher@wisc.edu

* The authors acknowledge financial support from NIA grants R01 AG060109 and the Center for Demography of Health and Aging (CDHA) at the University of Wisconsin–Madison under NIA core grant P30 AG17266. The authors thank Sherry Huhn-Gotzler for excellent editorial support.

Background

Sex differences in mortality have been documented for centuries, with an advantage in life expectancy for females existing essentially since the mid-18th 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. Indeed, a great deal of literature has been dedicated to looking at sex differences in mortality and survival. For instance, one paper by Bergeron-Boucher and colleagues (2020), found that the likelihood of males outliving females varies substantially across dozens of countries and regions since the middle of the 18th 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, researchers have observed a persistent female advantage, that has grown substantially over time. Due to economic development and improved living conditions for females, life expectancy increased considerably throughout much of the 20th century for women in most Western societies (Rigby & Dorling, 2007). For instance, Oksuzyan and colleagues (2008) found among Nordic countries that the gap between males and females was between 2–4 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. Other contexts have observed this pattern as well, in which 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 & 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, prior research has suggested that 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 & 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 19th 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. Reliable data on mortality only exists 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 19th century (Hacker, 2010; Pope, 1992). Ultimately, this prior research found that throughout the 19th century males and females largely possessed roughly equal life expectancy, with the former having a slight advantage. In an example of this, Goldin and Lleras-Muney (2019) show with data collected from the state of Massachusetts that younger-aged females (ages 5–25) were disproportionately affected by infectious disease, which undoubtedly contributed a male advantage in life expectancy. However, females began to gain an advantage in life expectancy during the

American Civil War as the conflict caused a large sudden increase in male mortality. In subsequent years, the female advantage was largely maintained. Other research also shows that birth cohorts 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 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

Studies on the female advantage as it relates to longevity have looked at it through a myriad of different lenses. Many of the explanations apportion the difference into a combination of biological differences and social and behavioral factors (Austad & Fischer, 2016; Crimmins et al., 2019; Rogers et al., 2010). Regarding the former, biologic differences in genetic structure and hormonal differences are viewed by some to favor female life expectancy (Austad, 2006). Research that supports this view from a biological standpoint has looked at populations with similar lifestyles in terms of health behaviors or religious practices (Brønnum-Hansen & Juel, 2001; Lindahl-Jacobsen et al., 2013). The findings indicate that in contexts where males and females have similar lifestyles, women still showcase a longevity advantage, implying a role for a biologic advantage. Even among other species in the animal kingdom, evidence indicates that females as a whole see an advantage relative to their male counterparts (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).

Turning to social explanations for differences in longevity, several studies have attributed the majority of the disparities to behavioral factors (Janssen, 2020; Rogers et al., 2010). First,

among males, their higher prevalence of risk-taking behaviors especially during younger years, puts them at risk to not survive to older ages. Second, males on average have higher consumption of tobacco[1] and alcohol, drive less safely, and eat less healthy than their female counterparts, which in turn heightens their risk for development of many chronic conditions and accidents (Beltrán-Sánchez et al., 2015).

In the context of the 19th 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 & 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 & Haines, 1991). Moreover, the majority of children who died before age 5 were males (Drevenstedt et al., 2008), but evidence shows that between the ages of 5 and 25, infectious disease disproportionately affected females (Goldin & Lleras-Muney, 2019).

Other scholars have shown evidence that sex differences in life expectancy throughout the century were due to adult mortality rates from specific illness declining more quickly among females. Beltrán-Sánchez and colleagues (2015) looked into birth cohorts from the 19th century and found that male-female mortality ratios in the latter half of the century increased for each subsequent cohort at ages 50, 60, 70, and 80. In the same research, they also found that among specific causes of death, increases in cardiovascular mortality among those age groups made the

[1] Furthermore, in developed nations, the greatest risk factor that contributes to differences in longevity is tobacco us (Preston & Wang, 2006).

largest contributions to these increased 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 19th century acquire 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 life span, meaning that lower parity essentially helps to give rise to longer lifespans (Ehrlich, 2015).. The results of these investigations often have conflicting results, some researchers find a tradeoff between parity and life expectancy (Smith et al., 2009) or potentially positive (McArdle et al., 2006; Muller et al., 2002) or negative correlations (Lycett et al., 2000)[2]. Altogether, a fuller understanding of the strength of this hypothesis likely requires much larger samples.

Finally, the role that geography plays in life expectancy has been thoroughly documented in the United States (Chetty et al., 2016; Deryugina & 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 which in turn can have an impact on mortality (Hayward et al., 1997). Furthermore, a great deal of research looks to examine demographic phenomena at smaller levels of geography, because broader levels of geography often mask disparities within. This is illustrated by Boing and colleages (2020), who looked specifically at geographic variation of life expectancy in the United States and found that smaller

[2] It 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.

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 19th century, a period where there is 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 & 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 paper seeks to examine how sex differences in longevity in the United States have changed over time, specifically across 19th -century birth cohorts. Specifically, this paper 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 state of birth; and 3) how parity shapes these differences. Ultimately, the aim of this paper is to increase our understanding of the dramatic societal change undergone historically with regard to the sex gap in longevity.

Methods

The data utilized in this research comes from genealogical data on the website Geni.com, which stores individual profiles that have been uploaded to family trees. 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.) is unavailable. Initially, this paper's authors

collected, cleaned, and later validated for use 86 million profiles from the site, with some dating as far back as the 17th 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 has been compared with the average lifespan from a worldwide historical analysis (Oeppen & Vaupel, 2002) to gauge its validity. Kaplanis and colleagues (2018) found alignments of R^2=95% between historical and Geni data, and a 98% concordance with those reported by the Human Mortality Database.

The Geni data source is unique as it provides a large and robust sample size, which is essential to examine trends in longevity among birth cohorts in the 19th century. Additionally, its large sample sizes allow for the examination of state-level differences. Other datasets that include similar information laid out in this paper 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). Despite these clear advantages, there some possible pitfalls to the data source, namely in the form of self-selection biases. For instance, due to the fact that the data is based off genealogical data and the fact that we opt to study the role of fertility, it is critical to note at the outset that the data may be biased towards those with higher numbers of children surviving to reproductive ages. This, along with other potential limitations that will be discussed later, are important to consider moving forward.

Starting from the initial 86 million individuals in the dataset, we dropped specific observations—those who were not born in the United States, or for those whose birth

information was missing—as the focus of this paper is on birth cohorts in the United States. We then excluded those who were born prior to 1800 and those born after 1900. These parameters allowed us to focus on individuals born in the 19th century, while also capturing the full mortality of birth cohorts by limiting issues of right censoring at 2015. Finally, we excluded observations for those individuals missing sex information and those with unreliable lifespan lengths (>110 years). The final analytic sample that was used for analysis was 1,394,499. See Table SA in the appendix for documentation on the sample size reductions.

The key outcome looked at in this paper is longevity, which is ascertained by date of birth and date of death. Due to the role that child mortality may play in estimates, all longevity estimates in this paper are conditional on the individual surviving to age 5. As a sensitivity check, we ran estimates conditional on survival to age 10. To account for geographic disparities that exist, we control for state of birth, which is recorded in the Geni data.

With regards to fertility, we control for the number of children that an individual has, both conditional on survival to age 15 and age 50. The inclusion of the latter was to capture the window of childbearing. Specifically, we code the number of children ranging from having one child to having five or more. It should be noted that when exploring longevity among those who had children, our sample is reduced to about a quarter of the size of the full sample, which is due in large part to being unable to discern missing information on children as indicative of not having children or not. While significant, this reduced sample still leaves us with several hundreds of observations from which to use. Moving on, we next opt to calculate lifespan variation from our longevity estimates, because they can both reflect disparities between groups at the population level, and see if females during this time period had lower variation like in modern times (Aburto et al., 2020; Edwards & Tuljapurkar, 2005). Finally, we calculate period

life expectancy from the Geni data to contrast with our cohort data using standard life table methods. Incorporating period estimates, which summarizes mortality risks experienced by different cohorts during one period of time, along with cohort estimates that measure the mortality experience of a cohort as they age over time from birth until the death of the final survivor (Canudas-Romo & Schoen, 2005), allow us to test the timing of differences and trends.

First, to address potential validity concerns, we opt to show how our data matches up against cohort estimates from the Human Mortality Database (HMD). Next, we present our cohort longevity estimates that we calculated from our massive individual-level dataset. Supplementary figures are presented regarding our period estimates of life expectancy, along with a comparison of our cohort and period estimates to highlight slight differences. Other figures focus specifically on models that include survival to age 5 and 10, while also controlling for place using state of birth. We then show trends in lifespan variation to highlight the underlying dispersion that exists in longevity among males and females. Finally, we turn to the number of children an individual has, first for the whole sample and then subsequently by gender, in models that are conditional on survival first to age 15 and then age 50, to account for the window of childbearing.

Results

Figure 1 displays our cohort estimates from 1800 to 1900 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 mirrors 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 paper. Ideally, we would have used younger age

cutoffs to show these trends, but due to the United States not having as far-reaching demographic data as other nations (i.e. Sweden), we opted to use older age cutoffs to add validity to our estimates before moving forward.

Table 1 displays the descriptive statistics of our sample. Among the 1,394,499 individuals in our sample, 757,941 (54.3%) were males and 636,558 (45.7%) were females. Due to conditioning on survival to age 5, our longevity estimates ranged from 5 to 110, with the average in our sample being 69.84 years. In terms of number of children, the average number of children an individual had in our sample was 3.17 children, and the average lifespan variation was 22.53 years. For sex-specific statistics, males had a slightly lower average longevity of 69.02, whereas females' average longevity was 70.81. However, these are average estimates for the entire sample; therefore, we next delve into trends across the birth decades by gender.

For our cohort life expectancy estimates, Figure 2 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, by 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 to 1900.

Regarding our period estimates of longevity, we note similar trends to our cohort estimates, with some estimates differing. To display these differences, we compare cohort and

period estimates in Figure 1A (see Appendix). Shown together, the period estimates reveal a persistent male advantage throughout the century, converging around the mid-point of the century. These differences are not entirely unexpected, given typical gaps and lags between period and cohort measures of longevity (Canudas-Romo & Schoen, 2005).

Figures 3 and 4 show results from models that regress longevity on birth decades from 1800 to 1900, with 95% confidence intervals. The former of the figures looks at our baseline estimates and the latter controls for 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 acceleration in longevity in later years. Interestingly, there is little change in estimates when the control for state of birth is incorporated, despite some states showcasing an association with longevity. As a robustness check, we again estimated the above models, but this time we look at longevity conditional on survival to age 10, which show similar estimates (see Appendix, Figures 2A and 3A, respectively).

Moving on to trends in the dispersion of longevity, we display lifespan variation estimates in Figure 5. 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 there begins a decline, 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 19th 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 Appendix, Figure 4A). Ultimately, we found consistent but narrowed gaps.

With regard to the role that childbearing has on longevity, Figure 6 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 is a great deal of variability in the longevity estimates, which begins to collapse around 1860, and then accelerates. To capture the window of childbearing, in Figure 7 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.

Figures 8 and 9 show the trends presented in Figure 6, but look at male and female longevity, respectively. Note that there is a great deal of variation for males and females at the start of the century. Furthermore, like the whole population, there is a collapse in this variation in the latter half of the century, but it is far greater for females. Additionally, the male longevity gains across the time periods are relatively stagnant or stable when controlling for number of children, whereas for females there is a steep increase of nearly fifteen years in some estimates. Next, Figures 10 and 11, show male and female longevity estimates varied by number of children, conditional on survival to age 50. Conditioning on surviving to 50 years of age does relatively little for males with regard to longevity, whereas it makes a large difference for females. That is, we see higher estimates in longevity and relatively little variation compared to Figure 10 where it is conditioned on surviving to age 10. Thus, regarding the overall trends,

these results indirectly support a mechanism in which the influence of fertility plays a role in sex differences in longevity.

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 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 death selection. 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), suggesting 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 19th century, implying that the trends remain largely unexplained.

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:

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

$\bar{Y}_B$ 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 older cohorts born before 1870, where we differentiate the endowment (i.e., the effects of decreases in fertility levels $\overline{X_{Aj}} - \overline{X_{Bj}}$, 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. To limit our attention to only the role of fertility, decomposition was done after controlling for state dummies.

First, decomposition in Table 1A (See Appendix) shows that the endowment effects are negative, which became smaller and negligible when conditioning on survival to 50, especially among females. Second, the coefficient effects are close to zero and not statistically significant. Third, the interaction effects, which are the changes in the level and coefficients of fertility taken together, are statistically significant among all groups, including those who survive to 50 years of age and especially among females. This implies that the joint changes in the distribution and effects of fertility partly explain the female time trends of longevity by a factor of about 9%. In sum, this 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

Previous research has made significant contributions to the literature surrounding sex differences in longevity, with females showing a clear advantage over males (Austad, 2006;

Barford et al., 2006). However, very few contributions to this literature in the context of the United States discuss differences in the 19th century, during which a slight male advantage existed (Goldin & 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 ascertain how sex differences in longevity in the United States changed across 19th -century birth cohorts. By looking at baseline differences by gender, along with incorporating geographic and fertility information, we found evidence that males did in fact hold an advantage over females until mid-century, after which females gained a dramatic advantage.

Using unique and massive individual data from our genealogic data source, the results in this paper 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 (Goldin & Lleras-Muney, 2019). This reinforces the notion that excess mortality among males during that time allowed females to gain an initial advantage. Second, this paper revealed some evidence of spatial variations in these patterns, in the form of state of birth. However, the introduction of state of birth in analysis did not attenuate the estimates of birth decade on longevity a great deal, as evidenced in our figures.

Third, our results revealed a novel finding in that females had higher lifespan variation throughout the entirety of the 19th century. This finding is largely inconsistent with prior literature, which finds that females have lower lifespan variation than males (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.

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 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 covers long historical periods and provides a promising avenue for demographers to use (Alburez-Gutierrez et al., 2019), the longevity data come from a select sample, and thus may have longer lifespans than the general population at that time.

Another limitation of this work is that Geni data does 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 on fertility may be underestimates 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. However, given that there was no method of differentiation regarding this, we opted to focus on fertility of those who did have children. Future research should aim to consider fertility of populations during this time to see if childlessness improved longevity or if the conclusions we reached with number of children are robust.

A third limitation of our study stems from a lack of data on race and 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 life spans, which allows us to not be as reliant on period measures, while also being able to factor in controls for geography and fertility.

Conclusion

The rise in life expectancy over the centuries signifies a remarkable feat in human history (Oeppen & Vaupel, 2002). This study is one of the first to utilize cohort rather than period measures of longevity in order to study a fundamental, yet recent, demographic phenomenon: the female advantage in life expectancy. Furthermore, this study is one of a select few that look at

this in the context of the United States. Ultimately, our study reveals that prior to the mid-18th century, males held a slight advantage over females, after which females gained and maintained this advantage. Our usage of massive individual-level genealogical data enabled us to analyze over a long historical period, while also controlling for factors such as place and fertility. In short, this research helps to pinpoint the importance of specific factors outside of historical processes that helped contribute to sex differences in survival.

References

Abrams, L. R., Myrskylä, M., & Mehta, N. K. (2021). The growing rural–urban divide in US life expectancy: Contribution of cardiovascular disease and other major causes of death. *International Journal of Epidemiology*, dyab158. https://doi.org/10.1093/ije/dyab158

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*(10), 5250–5259. https://doi.org/10.1073/pnas.1915884117

Alburez-Gutierrez, D., Zagheni, E., Aref, S., Gil-Clavel, S., Grow, A., & Negraia, D. V. (2019). *Demography in the Digital Era: New Data Sources for Population Research* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/24jp7

Austad, S. 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

Austad, S. N., & Fischer, K. E. (2016). Sex Differences in Lifespan. *Cell Metabolism*, *23*(6), 1022–1033. https://doi.org/10.1016/j.cmet.2016.05.019

Barford, A., Dorling, D., Smith, G. D., & Shaw, M. (2006). Life expectancy: Women now on top everywhere. *BMJ*, *332*(7545), 808. https://doi.org/10.1136/bmj.332.7545.808

Beltrán-Sánchez, H., Finch, C. E., & Crimmins, E. M. (2015). Twentieth century surge of excess adult male mortality. *Proceedings of the National Academy of Sciences*, *112*(29), 8993–8998. https://doi.org/10.1073/pnas.1421942112

Bergeron-Boucher, M.-P., Alvarez, J.-A., Kashnitsky, I., Zarulli, V., & Vaupel, J. W. (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

Boing, A. F., Boing, A. C., Cordes, J., Kim, R., & Subramanian, S. V. (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), 17688–17694. https://doi.org/10.1073/pnas.2003719117

Brønnum-Hansen, H., & Juel, K. (2001). Abstention 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), 273–278. https://doi.org/10.1136/tc.10.3.273

Canudas-Romo, V., & Schoen, R. (2005). Age-specific contributions to changes in the period and cohort life expectancy. *Demographic Research*, *13*, 63–82. https://doi.org/10.4054/DemRes.2005.13.3

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (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

Chiu, C.-T. (2019). Living arrangements and disability-free life expectancy in the United States. *PLOS ONE*, *14*(2), e0211894. https://doi.org/10.1371/journal.pone.0211894

Colchero, F., Rau, R., Jones, O. R., Barthold, J. A., Conde, D. A., Lenart, A., Nemeth, L., Scheuerlein, A., Schoeley, J., Torres, C., Zarulli, V., Altmann, J., Brockman, D. K., Bronikowski, A. M., Fedigan, L. M., Pusey, A. E., Stoinski, T. S., Strier, K. B., Baudisch, A., … Vaupel, J. W. (2016). The emergence of longevous populations. *Proceedings of the National Academy of Sciences*, *113*(48), E7681–E7690. https://doi.org/10.1073/pnas.1612191113

Condran, G. A., & Crimmins-Gardner, E. (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, E. M., Shim, H., Zhang, Y. S., & Kim, J. K. (2019). Differences between Men and Women in Mortality and the Health Dimensions of the Morbidity Process. *Clinical Chemistry*, *65*(1), 135–145. https://doi.org/10.1373/clinchem.2018.288332

Deryugina, T., & Molitor, D. (2021). The Causal Effects of Place on Health and Longevity. *Journal of Economic Perspectives*, *35*(4), 147–170. https://doi.org/10.1257/jep.35.4.147

Drevenstedt, G. L., Crimmins, E. M., Vasunilashorn, S., & Finch, C. E. (2008). The rise and fall of excess male infant mortality. *Proceedings of the National Academy of Sciences*, *105*(13), 5016–5021. https://doi.org/10.1073/pnas.0800221105

Dribe, M. (2004). Long-term effects of childbearing on mortality: Evidence from pre-industrial Sweden. *Population Studies*, *58*(3), 297–310. https://doi.org/10.1080/0032472042000272357

Dribe, M., Juárez, S. P., & Scalone, F. (2017). Is It Who You Are or Where You Live? Community Effects on Net Fertility at the Onset of Fertility Decline: A Multilevel Analysis Using Swedish Micro-Census Data: Is It Who You Are or Where You Live? *Population, Space and Place*, *23*(2), e1987. https://doi.org/10.1002/psp.1987

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*(4), 645–674. https://doi.org/10.1111/j.1728-4457.2005.00092.x

Ehrlich, S. (2015). Effect of fertility and infertility on longevity. *Fertility and Sterility*, *103*(5), 1129–1135. https://doi.org/10.1016/j.fertnstert.2015.03.021

Goldin, C., & Lleras-Muney, A. (2019). XX > XY?: The changing female advantage in life expectancy. *Journal of Health Economics*, *67*, 102224. https://doi.org/10.1016/j.jhealeco.2019.102224

Hacker, J. D. (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, M. (2001). *The Urban Mortality Transition in the United States, 1800-1940* (No. h0134; p. h0134). National Bureau of Economic Research. https://doi.org/10.3386/h0134

Hayward, M. D., Pienta, A. M., & McLaughlin, D. K. (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

Janssen, F. (2020). Changing contribution of smoking to the sex differences in life expectancy in Europe, 1950–2014. *European Journal of Epidemiology*, *35*(9), 835–841. https://doi.org/10.1007/s10654-020-00602-x

Kalben, B. B. (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, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovits, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D. G., Price, A. L., & Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*, *360*(6385), 171–175. https://doi.org/10.1126/science.aam9309

Kirk, D. (1996). Demographic Transition Theory. *Population Studies*, *50*(3), 361–387. https://doi.org/10.1080/0032472031000149536

Lindahl-Jacobsen, R., Hanson, H. A., Oksuzyan, A., Mineau, G. P., Christensen, K., & Smith, K. R. (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), 161–166. https://doi.org/10.1016/j.annepidem.2013.02.001

Luy, M., & Gast, K. (2014). Do Women Live Longer or Do Men Die Earlier? Reflections on the Causes of Sex Differences in Life Expectancy. *Gerontology*, *60*(2), 143–153. https://doi.org/10.1159/000355310

Lycett, J. E., Dunbar, R. I., & Voland, E. (2000). Longevity and the costs of reproduction in a historical human population. *Proceedings. Biological Sciences*, *267*(1438), 31–35. https://doi.org/10.1098/rspb.2000.0962

McArdle, P. F., Pollin, T. I., O'Connell, J. R., Sorkin, J. D., Agarwala, R., Schäffer, A. A., Streeten, E. A., King, T. M., Shuldiner, A. R., & Mitchell, B. D. (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), 190–195. https://doi.org/10.1093/gerona/61.2.190

Modig, K., Andersson, T., Vaupel, J., Rau, R., & Ahlbom, A. (2017). How long do centenarians survive? Life expectancy and maximum lifespan. *Journal of Internal Medicine*, *282*(2), 156–163. https://doi.org/10.1111/joim.12627

Montez, J. K., Beckfield, J., Cooney, J. K., Grumbach, J. M., Hayward, M. D., Koytak, H. Z., Woolf, S. H., & Zajacova, A. (2020). US State Policies, Politics, and Life Expectancy. *The Milbank Quarterly*, *98*(3), 668–699. https://doi.org/10.1111/1468-0009.12469

Mosca, L., Barrett-Connor, E., & Kass Wenger, N. (2011). Sex/Gender Differences in Cardiovascular Disease Prevention: What a Difference a Decade Makes. *Circulation*, *124*(19), 2145–2154. https://doi.org/10.1161/CIRCULATIONAHA.110.968792

Muller, H.-G., Chiou, J.-M., Carey, J. R., & Wang, J.-L. (2002). Fertility and Life Span: Late Children Enhance Female Longevity. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *57*(5), B202–B206. https://doi.org/10.1093/gerona/57.5.B202

Oeppen, J., & Vaupel, J. W. (2002). Broken Limits to Life Expectancy. *Science*, *296*(5570), 1029–1031. https://doi.org/10.1126/science.1069675

Oksuzyan, A., Juel, K., Vaupel, J. W., & Christensen, K. (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

Omran, A. R. (2005). The Epidemiologic Transition: A Theory of the Epidemiology of Population Change: *The Epidemiologic Transition*. *Milbank Quarterly*, *83*(4), 731–757. https://doi.org/10.1111/j.1468-0009.2005.00398.x

Pope, C. L. (1992). Adult Mortality in America before 1900: A View from Family Histories. In *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel* (pp. 267–296). University of Chicago Press.

Preston, S. H., & Haines, M. R. (1991). *Fatal years: Child mortality in late nineteenth-century America*. Princeton University Press.

Preston, S. H., & Wang, H. (2006). Sex mortality differences in The United States: The role of cohort smoking patterns. *Demography*, *43*(4), 631–646. https://doi.org/10.1353/dem.2006.0037

Rigby, J. E., & Dorling, D. (2007). Mortality in relation to sex in the affluent world. *Journal of Epidemiology & Community Health*, *61*(2), 159–164. https://doi.org/10.1136/jech.2006.047381

Rogers, R. G., Everett, B. G., Onge, J. M. S., & Krueger, P. M. (2010). Social, behavioral, and biological factors, and sex differences in mortality. *Demography*, *47*(3), 555–578. https://doi.org/10.1353/dem.0.0119

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

Seifarth, J. E., McGowan, C. L., & Milne, K. J. (2012). Sex and Life Expectancy. *Gender Medicine*, *9*(6), 390–401. https://doi.org/10.1016/j.genm.2012.10.001

Singh, G. K., & Siahpush, M. (2014). Widening Rural–Urban Disparities in Life Expectancy, U.S., 1969–2009. *American Journal of Preventive Medicine*, *46*(2), e19–e29. https://doi.org/10.1016/j.amepre.2013.10.017

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

Sundberg, L., Agahi, N., Fritzell, J., & Fors, S. (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), 673–681. https://doi.org/10.1007/s00038-018-1097-3

Tavassoli, N., Noghanibehambari, H., Noghani, F., & Toranji, M. (2020). Upswing in Industrial Activity and Infant Mortality during Late 19th Century US. *Journal of Environments*, *6*(1), 1–13. https://doi.org/10.20448/journal.505.2020.61.1.13

Weidner, G. (2000). Why Do Men Get More Heart Disease Than Women? An International Perspective. *Journal of American College Health*, *48*(6), 291–294. https://doi.org/10.1080/07448480009596270

Zafeiris, K. 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, V., Barthold Jones, J. A., Oksuzyan, A., Lindahl-Jacobsen, R., Christensen, K., & Vaupel, J. W. (2018). Women live longer than men even during severe famines and epidemics. *Proceedings of the National Academy of Sciences*, *115*(4), E832–E840. https://doi.org/10.1073/pnas.1701535115

Appendix

Figure 1A. Period Life Expectancy, Conditional on Survival to Specific Ages

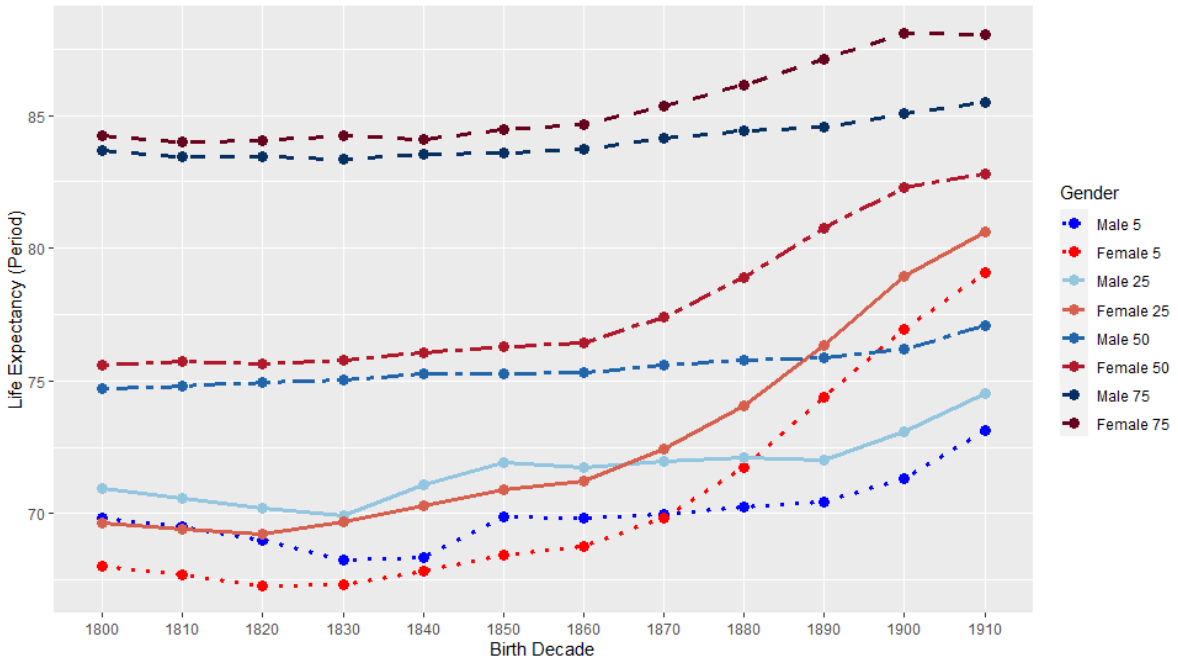

Figure 2A. Period Life Expectancy vs. Longevity Conditional on Survival to Age 5

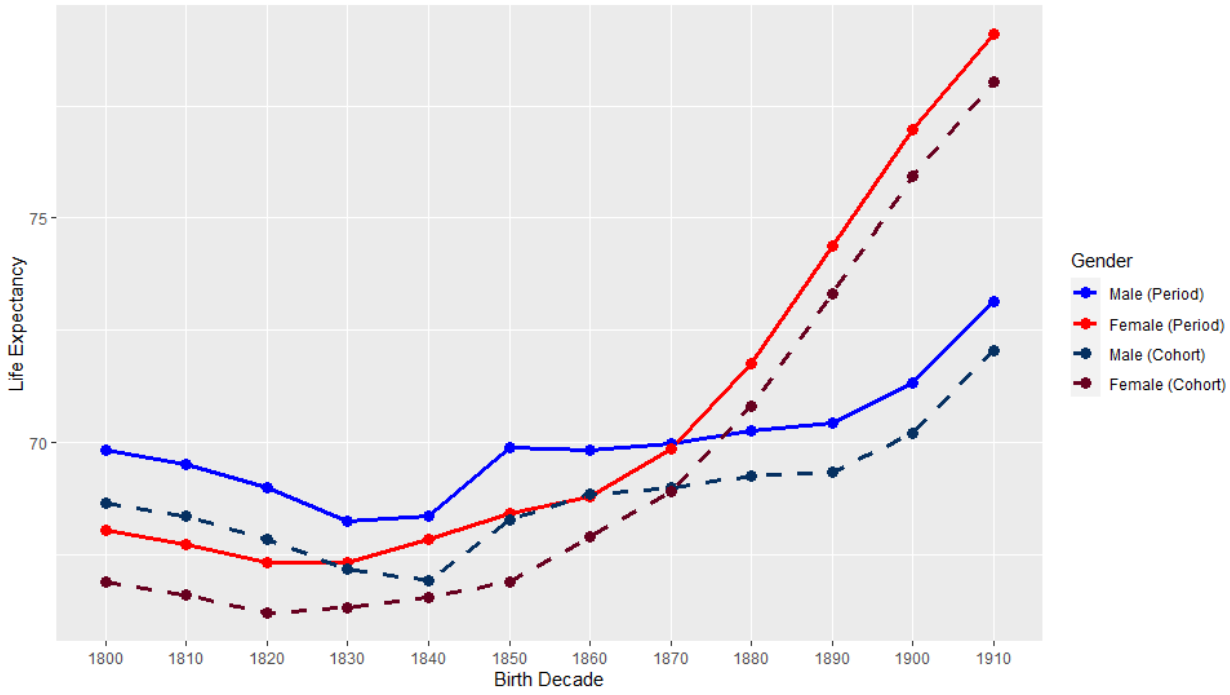

Figure 3A. Life Expectancy Conditional on Survival to Age 10

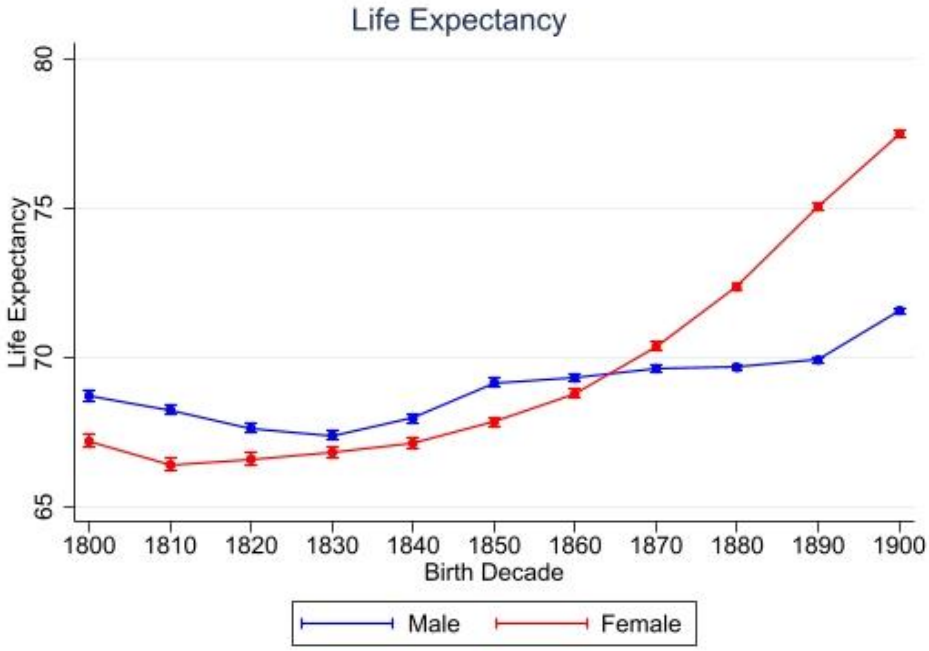

Figure 4A. Life Expectancy Conditional on Survival to Age 10 (Controlling for State of Birth)

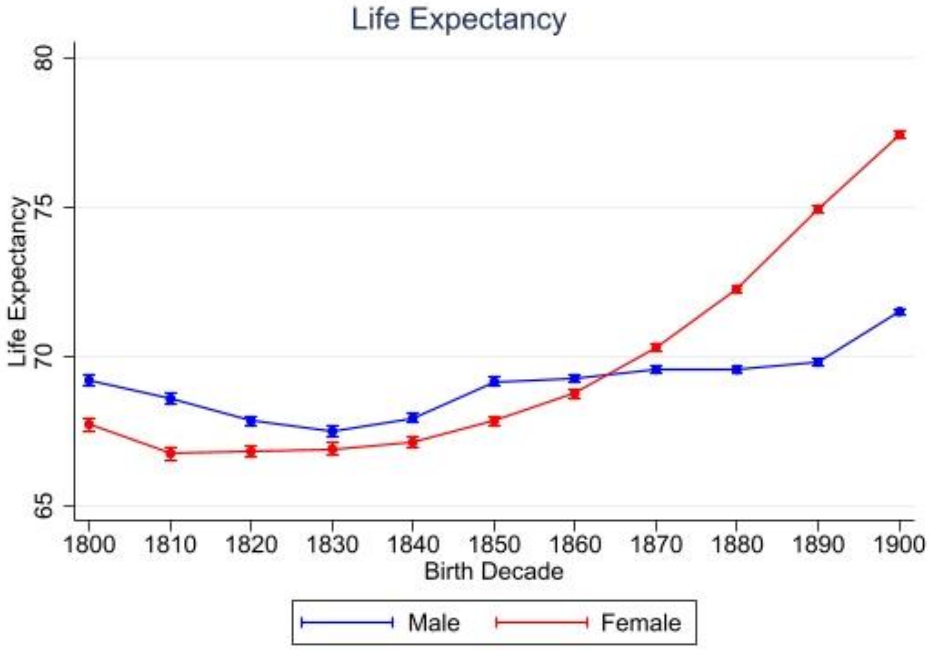

Figure 5A. Lifespan Variation Conditional on Survival to Age 50

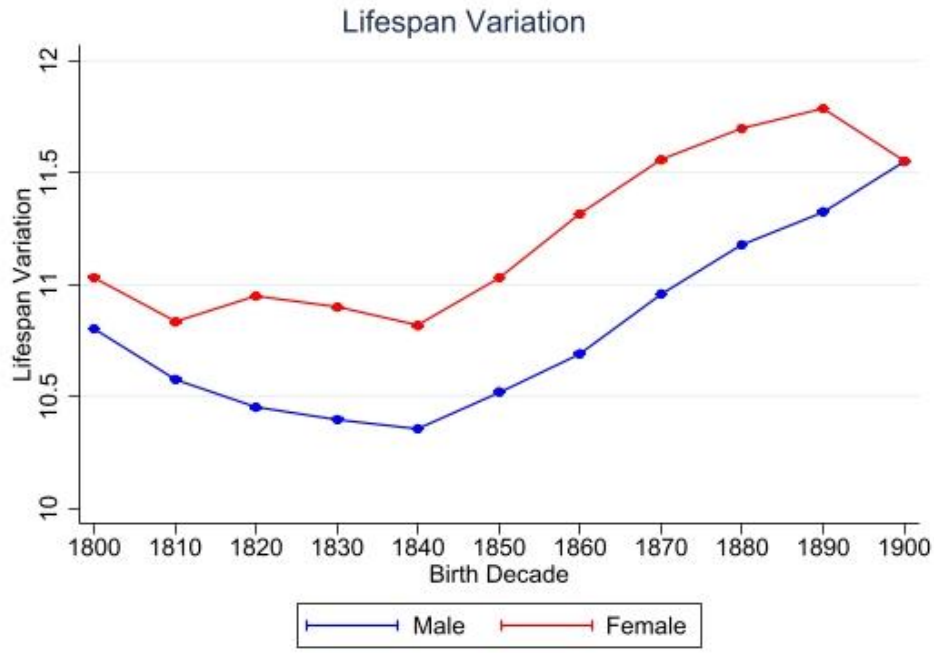

Table SA: Sample Selection

Initial sample	86,124,644
Valid birth & death dates	11,227,313
Longevity >=5 & <=110	10,001,726
Valid birthplace	5,500,020
Valid gender	5,499,386
Final sample (US born, 1800-1900)	1,394,499

Table 1A. Kitigawa-Blinder-Oaxaca Decomposition of Difference in Longevity

	(1)	(2)	(3)	(4)
Summary				
<1870	69.928***	71.837***	68.665***	72.696***
	(0.210)	(0.182)	(0.254)	(0.207)
≥1870	73.004***	74.290***	77.284***	79.270***
	(0.074)	(0.067)	(0.087)	(0.073)
Difference	3.076***	2.453***	8.619***	6.574***
	(0.223)	(0.194)	(0.268)	(0.219)
Decomposition				
Endowments	-0.865***	-0.290***	-0.895***	-0.037
	(0.034)	(0.030)	(0.042)	(0.036)
Coefficients	-0.049	-0.061+	-0.036	-0.042
	(0.038)	(0.033)	(0.047)	(0.037)
Interaction	0.835***	0.395***	1.299***	0.615***
	(0.045)	(0.041)	(0.053)	(0.046)
Intercept	3.155***	2.409***	8.251***	6.039***
	(0.228)	(0.198)	(0.274)	(0.224)
Gender	Male	Male	Female	Female
Age cutoff	5	40	5	40
N	199,796	192,484	166,537	156,543

Standard errors in parentheses *** p<0.001, ** p<0.01, * p<0.05

Table 2A. Sex Differences in Survival, Ages 5 and 10

	(1)	(2)	(3)	(4)
Sample	Age 5	Age 5	Age 10	Age 10
State of Birth	No	Yes	No	Yes
Birth Cohort (ref. 1800)				
1810	-0.498***	-0.696***	-0.469***	-0.636***
	(0.137)	(0.137)	(0.132)	(0.132)
1820	-1.149***	-1.491***	-1.086***	-1.374***
	(0.135)	(0.135)	(0.130)	(0.130)
1830	-1.515***	-1.971***	-1.323***	-1.711***
	(0.133)	(0.133)	(0.128)	(0.129)
1840	-1.072***	-1.644***	-0.770***	-1.261***
	(0.129)	(0.130)	(0.125)	(0.125)
1850	-0.0305	-0.576***	0.439***	-0.047
	(0.125)	(0.126)	(0.121)	(0.122)
1860	0.324***	-0.257**	0.588***	0.057
	(0.123)	(0.125)	(0.118)	(0.120)
1870	0.525***	-0.085	0.914***	0.352***
	(0.119)	(0.121)	(0.115)	(0.117)
1880	0.686***	0.049	0.936***	0.341***
	(0.116)	(0.119)	(0.112)	(0.114)
1890	1.028***	0.398***	1.190***	0.599***
	(0.114)	(0.117)	(0.110)	(0.112)
1900	2.703***	2.103***	2.844***	2.280***
	(0.113)	(0.115)	(0.109)	(0.111)
Female	-1.647***	-1.601***	-1.525***	-1.493***
	(0.154)	(0.153)	(0.148)	(0.148)
Female X 1810	-0.246	-0.268	-0.318	-0.334
	(0.211)	(0.211)	(0.204)	(0.203)
Female X 1820	0.506**	0.475**	0.498**	0.475**
	(0.207)	(0.206)	(0.199)	(0.199)
Female X 1830	0.923***	0.872***	0.948***	0.908***
	(0.203)	(0.202)	(0.195)	(0.195)
Female X 1840	0.726***	0.687***	0.705***	0.676***
	(0.197)	(0.196)	(0.190)	(0.189)
Female X 1850	0.297	0.261	0.199	0.170
	(0.190)	(0.189)	(0.183)	(0.183)
Female X 1860	1.066***	1.047***	1.002***	0.988***
	(0.186)	(0.186)	(0.179)	(0.179)
Female X 1870	2.323***	2.292***	2.274***	2.250***
	(0.181)	(0.181)	(0.175)	(0.174)
Female X 1880	4.287***	4.266***	4.220***	4.206***
	(0.177)	(0.177)	(0.171)	(0.170)
Female X 1890	6.704***	6.672***	6.655***	6.634***

	(0.174)	(0.174)	(0.168)	(0.167)
Female X 1900	7.536***	7.493***	7.452***	7.423***
	(0.172)	(0.172)	(0.166)	(0.165)
Constant	68.51***	62.988***	68.721***	63.440***
	(0.0997)	(0.607)	(0.096)	(0.587)
N	1,394,499	1,394,499	1,384,319	1,384,319
R-squared	0.025	0.029	0.027	0.030

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table 3A. Sex Differences in Lifespan Variation, Ages 5 and 40

Sample	(1) Age 5	(2) Age 40
Birth Cohort (ref. 1800)		
1810	0.170***	-0.226***
	(0.0168)	(0.002)
1820	1.462***	-0.351***
	(0.0165)	(0.002)
1830	2.716***	-0.405***
	(0.0163)	(0.002)
1840	3.223***	-0.446***
	(0.0158)	(0.002)
1850	3.331***	-0.281***
	(0.0153)	(0.002)
1860	3.544***	-0.111***
	(0.0150)	(0.002)
1870	3.889***	0.156***
	(0.0146)	(0.002)
1880	3.782***	0.376***
	(0.0142)	(0.002)
1890	3.447***	0.523***
	(0.0140)	(0.002)
1900	2.929***	0.746***
	(0.0138)	(0.002)
Female	2.1070**	0.228***
	(0.0188)	(0.003)
Female X 1810	0.154***	0.027***
	(0.0258)	(0.004)
Female X 1820	-0.462***	0.269***
	(0.0253)	(0.004)
Female X 1830	-0.952***	0.274***
	(0.0248)	(0.004)
Female X 1840	-0.875***	0.230***
	(0.0240)	(0.003)
Female X 1850	-0.548***	0.281***
	(0.0232)	(0.003)
Female X 1860	-0.550***	0.395***
	(0.0227)	(0.003)
Female X 1870	-0.203***	0.369***
	(0.0221)	(0.003)
Female X 1880	-0.410***	0.290***
	(0.0216)	(0.003)
Female X 1890	-0.296***	0.232***
	(0.0213)	(0.003)
Female X 1900	-0.851***	-0.225***
	(0.0210)	(0.003)

Constant	18.82***	10.804***
	(0.0121)	(0.002)
Observations	1,394,491	1,191,048
R-squared	0.257	0.686

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 4A. Predicted Longevity, Varied by Number of Children

Full Sample?	(1) Yes Age 15	(2) Yes Age 40	(3) Male Age 15	(4) Male Age 40	(5) Female Age 15	(6) Female Age 40
Birth Cohort (ref. 1800)						
1810	-0.879**	-0.289	-0.641	-0.122	-1.184**	-0.520
	(0.280)	(0.215)	(0.344)	(0.275)	(0.457)	(0.339)
1820	-0.562*	0.110	-1.002**	0.341	0.052	-0.112
	(0.287)	(0.220)	(0.354)	(0.284)	(0.466)	(0.342)
1830	-1.033***	-0.006	-1.403***	0.273	-0.498	-0.289
	(0.290)	(0.223)	(0.358)	(0.289)	(0.471)	(0.347)
1840	-0.0185	0.007	0.054	0.157	0.068	-0.064
	(0.290)	(0.221)	(0.362)	(0.288)	(0.466)	(0.342)
1850	0.0303	0.252	-0.236	0.169	0.491	0.459
	(0.280)	(0.214)	(0.348)	(0.278)	(0.451)	(0.331)
1860	0.321	0.574**	0.348	0.721**	0.527	0.561
	(0.280)	(0.214)	(0.351)	(0.280)	(0.447)	(0.328)
1870	1.580***	1.624***	1.024**	1.371***	2.410***	2.069***
	(0.273)	(0.208)	(0.341)	(0.271)	(0.438)	(0.320)
1880	2.219***	2.235***	1.106***	1.308***	3.679***	3.480***
	(0.266)	(0.203)	(0.332)	(0.264)	(0.428)	(0.313)
1890	4.396***	3.448***	1.718***	1.607***	7.811***	5.850***
	(0.255)	(0.193)	(0.315)	(0.250)	(0.414)	(0.300)
1900	6.321***	4.025***	3.085***	1.845***	10.201***	6.672***
	(0.241)	(0.182)	(0.300)	(0.237)	(0.390)	(0.281)
Number of Children (ref. 1 Child)						
2 Children	-3.512***	-0.558	-2.046***	-0.043	-5.371***	-1.360*
	(0.481)	(0.381)	(0.592)	(0.484)	(0.783)	(0.607)
3 Children	-1.499**	0.387	-1.549*	0.297	-1.530	0.452
	(0.508)	(0.397)	(0.616)	(0.500)	(0.845)	(0.638)
4 Children	-1.604**	-0.236	-1.431*	0.240	-1.889*	-0.940
	(0.509)	(0.396)	(0.621)	(0.505)	(0.839)	(0.627)
5 Children	1.350***	-0.086	1.197***	0.194	1.457***	-0.552
	(0.259)	(0.197)	(0.316)	(0.251)	(0.428)	(0.312)
2 Children X 1810	-1.512***	-0.086	0.282	-0.263	4.079***	1.527
	(0.451)	(0.354)	(0.807)	(0.659)	(1.084)	(0.830)
2 Children X 1820	-2.806***	-0.739*	-0.853	-0.722	2.683*	0.671
	(0.442)	(0.350)	(0.807)	(0.666)	(1.059)	(0.809)
2 Children X 1830	-3.182***	-0.701*	-1.012	-0.875	2.012	0.903
	(0.423)	(0.337)	(0.791)	(0.651)	(1.040)	(0.801)
2 Children X 1840	-2.636***	-0.558	-0.028	-0.742	1.854	0.969
	(0.407)	(0.317)	(0.779)	(0.633)	(1.022)	(0.785)
2 Children X 1850	-2.007***	-0.605*	0.787	-0.642	2.366*	0.788
	(0.376)	(0.292)	(0.755)	(0.613)	(0.989)	(0.758)
2 Children X 1860	-0.119	0.475	1.404	0.260	5.771***	2.090**

	(0.352)	(0.270)	(0.741)	(0.601)	(0.960)	(0.732)
2 Children X 1870	-0.0308	-0.063	1.932**	-0.342	5.393***	1.673*
	(0.321)	(0.244)	(0.717)	(0.580)	(0.935)	(0.712)
2 Children X 1880	0.921**	0.206	1.789**	-0.226	7.748***	2.165**
	(0.288)	(0.217)	(0.694)	(0.562)	(0.907)	(0.690)
2 Children X 1890	0.744**	0.316	2.677***	0.203	6.410***	1.998**
	(0.245)	(0.183)	(0.665)	(0.539)	(0.879)	(0.669)
2 Children X 1900	1.114***	0.731***	2.922***	0.691	7.040***	2.364***
	(0.202)	(0.148)	(0.644)	(0.522)	(0.846)	(0.646)
3 Children X 1810	-0.889	0.402	0.785	0.108	0.258	-0.178
	(0.474)	(0.371)	(0.840)	(0.684)	(1.160)	(0.876)
3 Children X 1820	-3.124***	-0.822*	-1.548	-1.536*	-1.708	-0.801
	(0.464)	(0.368)	(0.837)	(0.687)	(1.140)	(0.863)
3 Children X 1830	-1.468***	-0.436	0.326	-1.004	-0.167	-0.504
	(0.436)	(0.341)	(0.824)	(0.672)	(1.093)	(0.823)
3 Children X 1840	-1.508***	-0.495	0.047	-0.924	-0.012	-0.803
	(0.432)	(0.334)	(0.822)	(0.666)	(1.088)	(0.817)
3 Children X 1850	-1.049**	-0.203	1.599*	-0.203	-1.004	-1.169
	(0.392)	(0.302)	(0.785)	(0.634)	(1.055)	(0.794)
3 Children X 1860	0.371	0.276	1.857*	-0.153	1.782	-0.148
	(0.367)	(0.280)	(0.768)	(0.620)	(1.030)	(0.770)
3 Children X 1870	0.122	-0.323	1.420	-1.206*	1.915	-0.060
	(0.332)	(0.251)	(0.744)	(0.600)	(0.996)	(0.744)
3 Children X 1880	0.708*	0.078	1.616*	-0.451	3.022**	-0.067
	(0.299)	(0.226)	(0.722)	(0.583)	(0.970)	(0.724)
3 Children X 1890	1.015***	0.410*	2.419***	-0.005	2.638**	0.037
	(0.260)	(0.194)	(0.696)	(0.562)	(0.942)	(0.702)
3 Children X 1900	1.146***	0.488**	2.716***	0.233	2.915**	0.179
	(0.220)	(0.161)	(0.675)	(0.544)	(0.915)	(0.683)
4 Children X 1810	-0.730	0.331	0.488	0.147	1.356	1.156
	(0.476)	(0.372)	(0.849)	(0.693)	(1.151)	(0.861)
4 Children X 1820	-1.635***	-0.915**	0.402	-0.919	-0.664	-0.448
	(0.453)	(0.353)	(0.831)	(0.678)	(1.123)	(0.839)
4 Children X 1830	0.521	0.039	2.170**	-0.326	2.084	1.067
	(0.450)	(0.346)	(0.832)	(0.674)	(1.113)	(0.828)
4 Children X 1840	-0.680	-0.825*	0.873	-0.988	0.957	-0.065
	(0.425)	(0.325)	(0.816)	(0.658)	(1.081)	(0.802)
4 Children X 1850	0.245	0.244	2.024*	-0.078	1.641	1.237
	(0.404)	(0.309)	(0.799)	(0.645)	(1.061)	(0.789)
4 Children X 1860	1.021**	0.250	1.960*	-0.289	3.399***	1.499*
	(0.377)	(0.287)	(0.782)	(0.631)	(1.029)	(0.762)
4 Children X 1870	0.876*	-0.022	2.553***	-0.188	2.477*	0.814
	(0.353)	(0.267)	(0.765)	(0.616)	(1.006)	(0.745)
4 Children X 1880	1.682***	0.087	2.859***	0.119	3.870***	0.702
	(0.324)	(0.243)	(0.743)	(0.599)	(0.983)	(0.726)
4 Children X 1890	1.051***	0.063	2.554***	-0.112	2.743**	0.850

	(0.289)	(0.214)	(0.718)	(0.579)	(0.957)	(0.706)
4 Children X 1900	0.674**	0.215	2.578***	0.318	2.206*	0.913
	(0.251)	(0.183)	(0.695)	(0.560)	(0.930)	(0.686)
5+ Children X 1810	2.067***	0.323	0.611	0.280	0.880	0.623
	(0.262)	(0.201)	(0.451)	(0.358)	(0.608)	(0.445)
5+ Children X 1820	1.710***	-0.066	0.771	-0.063	-0.155	0.109
	(0.271)	(0.207)	(0.460)	(0.367)	(0.615)	(0.448)
5+ Children X 1830	2.746***	0.152	1.991***	0.127	0.773	0.486
	(0.275)	(0.210)	(0.466)	(0.371)	(0.615)	(0.449)
5+ Children X 1840	2.451***	0.428*	1.431**	0.507	0.726	0.582
	(0.271)	(0.204)	(0.465)	(0.367)	(0.606)	(0.441)
5+ Children X 1850	2.371***	0.131	1.619***	0.431	0.419	0.081
	(0.253)	(0.191)	(0.449)	(0.355)	(0.586)	(0.427)
5+ Children X 1860	2.852***	0.350	1.473**	0.355	1.601**	0.688
	(0.253)	(0.191)	(0.452)	(0.358)	(0.581)	(0.422)
5+ Children X 1870	2.057***	-0.146	0.916*	-0.077	0.564	0.123
	(0.244)	(0.184)	(0.444)	(0.351)	(0.574)	(0.416)
5+ Children X 1880	2.524***	0.237	1.262**	0.227	1.137*	0.559
	(0.235)	(0.176)	(0.435)	(0.343)	(0.564)	(0.408)
5+ Children X 1890	1.407***	-0.100	0.342	-0.080	-0.488	0.005
	(0.223)	(0.166)	(0.424)	(0.334)	(0.555)	(0.399)
5+ Children X 1900	0.769***	-0.205	-0.140	-0.039	-1.085*	-0.099
	(0.210)	(0.154)	(0.414)	(0.325)	(0.540)	(0.386)
Female	1.208***	2.982***				
	(0.0588)	(0.044)				
Constant	68.73***	72.994***	69.823***	73.612***	68.495***	75.100***
	(0.196)	(0.150)	(0.239)	(0.190)	(0.319)	(0.235)
Observations	290,279	258,759	158,879	143,809	131,400	114,950
R-squared	0.035	0.041	0.017	0.006	0.062	0.063

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 1. Descriptive Statistics

	All (N=1,394,499)		Male (N=757,941)		Female (N=636,558)	
Variable	Mean (SD)	Min, Max	Mean (SD)	Min, Max	Mean (SD)	Min, Max
Longevity	69.84 (18.81)	(5, 110)	69.02 (17.84)	(5, 110)	70.81 (20.07)	(5, 110)
Birth Decade						
1800	0.043	(0, 1)	0.046	(0, 1)	0.040	(0, 1)
1810	0.049	(0, 1)	0.052	(0, 1)	0.045	(0, 1)
1820	0.053	(0, 1)	0.056	(0, 1)	0.050	(0, 1)
1830	0.058	(0, 1)	0.059	(0, 1)	0.056	(0, 1)
1840	0.067	(0, 1)	0.068	(0, 1)	0.066	(0, 1)
1850	0.081	(0, 1)	0.081	(0, 1)	0.081	(0, 1)
1860	0.091	(0, 1)	0.089	(0, 1)	0.093	(0, 1)
1870	0.110	(0, 1)	0.108	(0, 1)	0.111	(0, 1)
1880	0.130	(0, 1)	0.128	(0, 1)	0.132	(0, 1)
1890	0.151	(0, 1)	0.149	(0, 1)	0.152	(0, 1)
1900	0.167	(0, 1)	0.164	(0, 1)	0.171	(0, 1)
Number of Children	3.17 (1.66)	(1, 5)	3.16 (1.65)	(1, 5)	3.18 (1.67)	(1, 5)
Lifespan Variation	22.53 (2.65)	(0, 41.42)	21.77 (2.59)	(2.81, 37.12)	23.44 (2.42)	(0, 41.42)

Figure 1. Geni Data Compared with Human Mortality Database Estimates

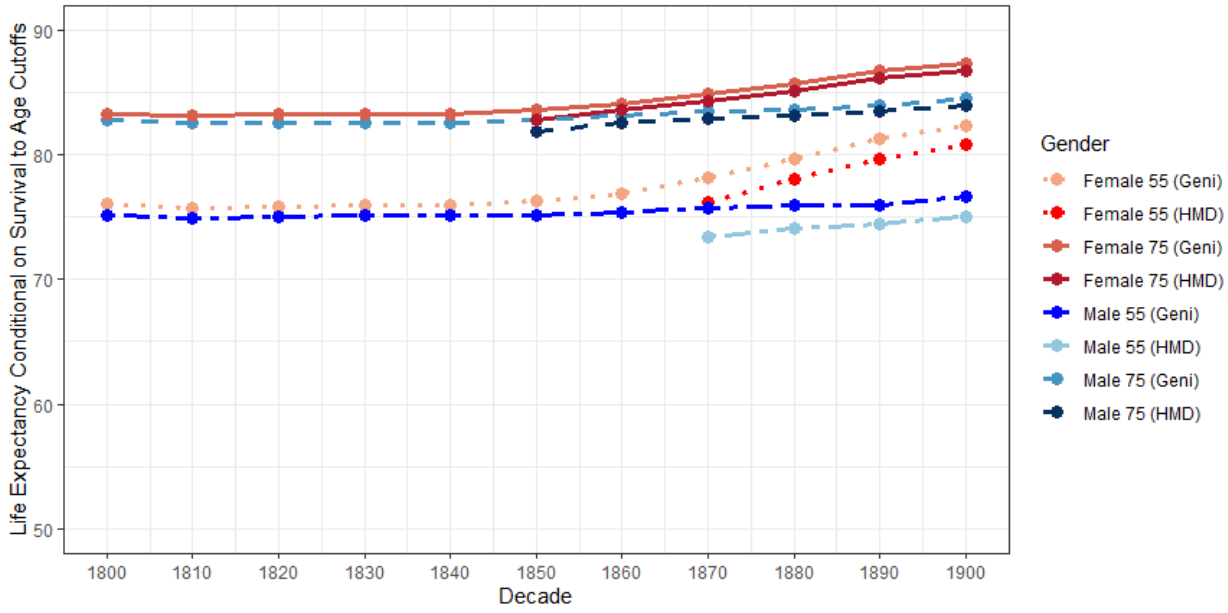

Figure 2. Predicted Life Expectancy Conditional on Survival to Age 5

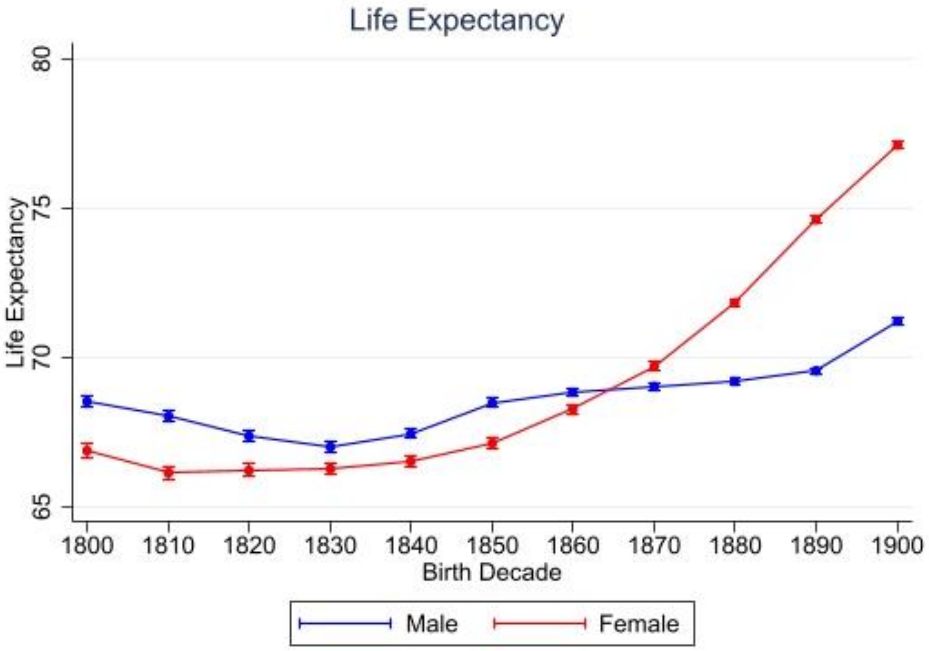

Figure 3. Predicted Life Expectancy Conditional on Survival to Age 5 (Controlling for State of Birth)

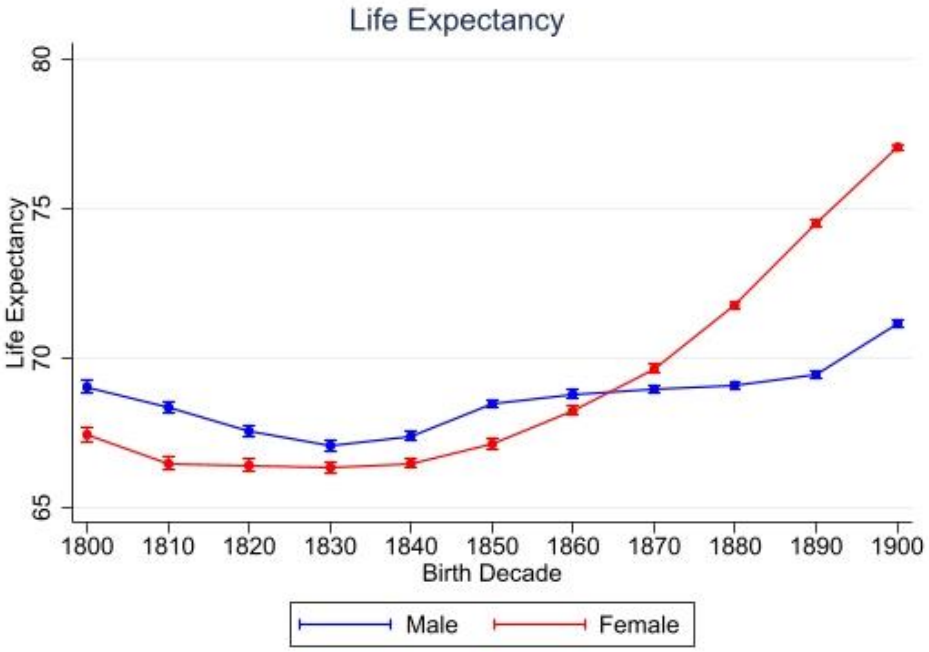

Figure 4. Lifespan Variation Conditional on Survival to Age 5

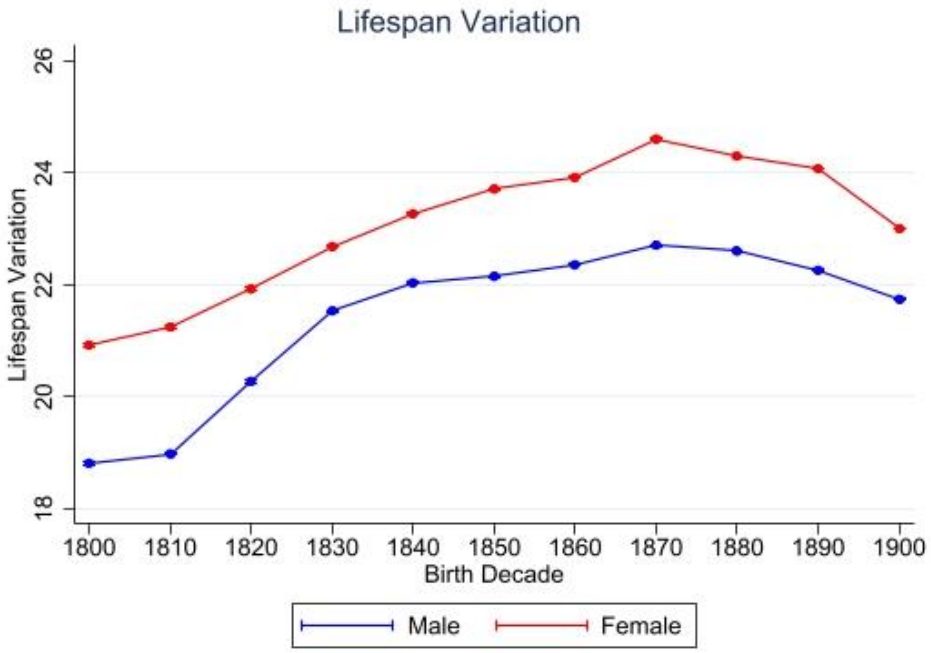

Figure 5. Predicted Life Expectancy at Age 15 by Number of Children

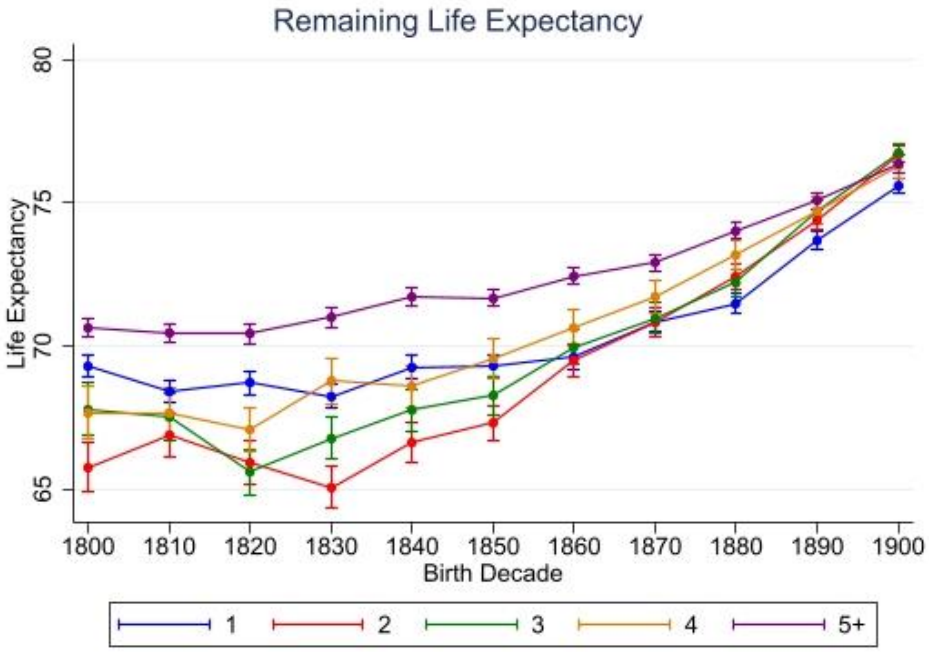

Figure 6. Predicted Life Expectancy at Age 50 by Number of Children

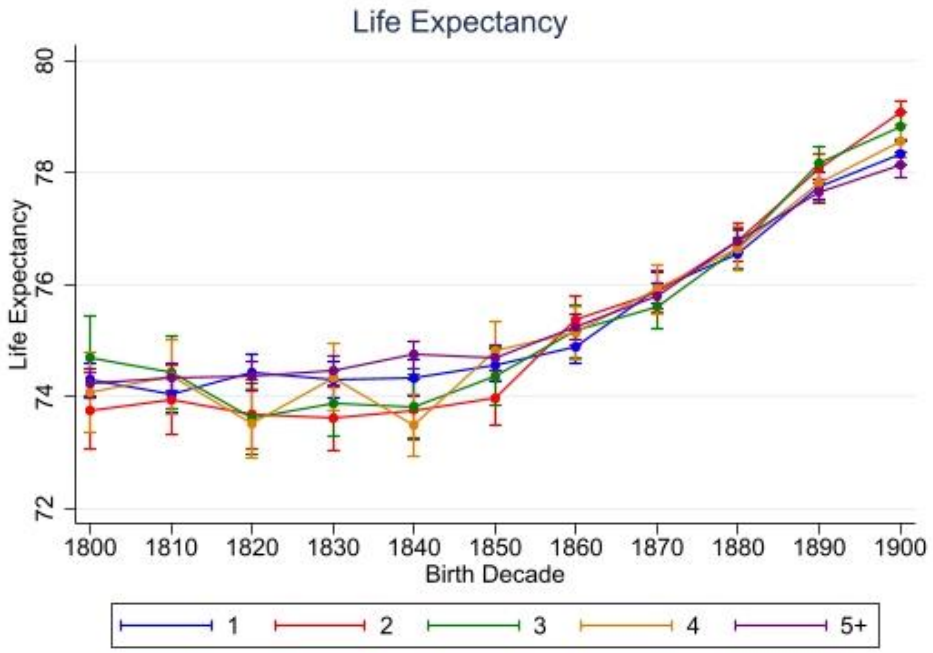

Figure 7. Predicted Life Expectancy at Age 15 by Number of Children, Males

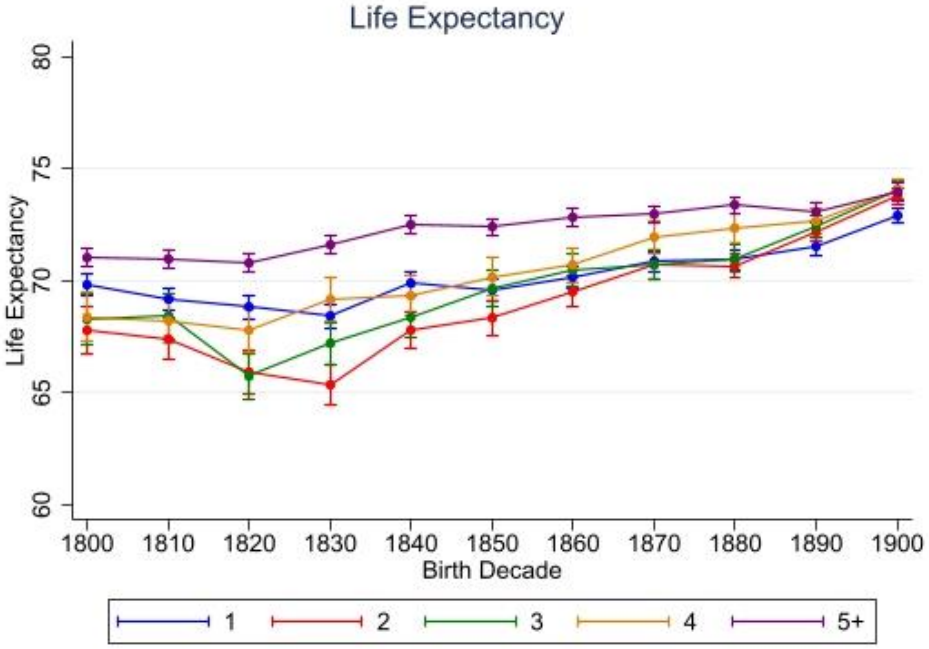

Figure 8. Predicted Life Expectancy at Age 15 by Number of Children, Females

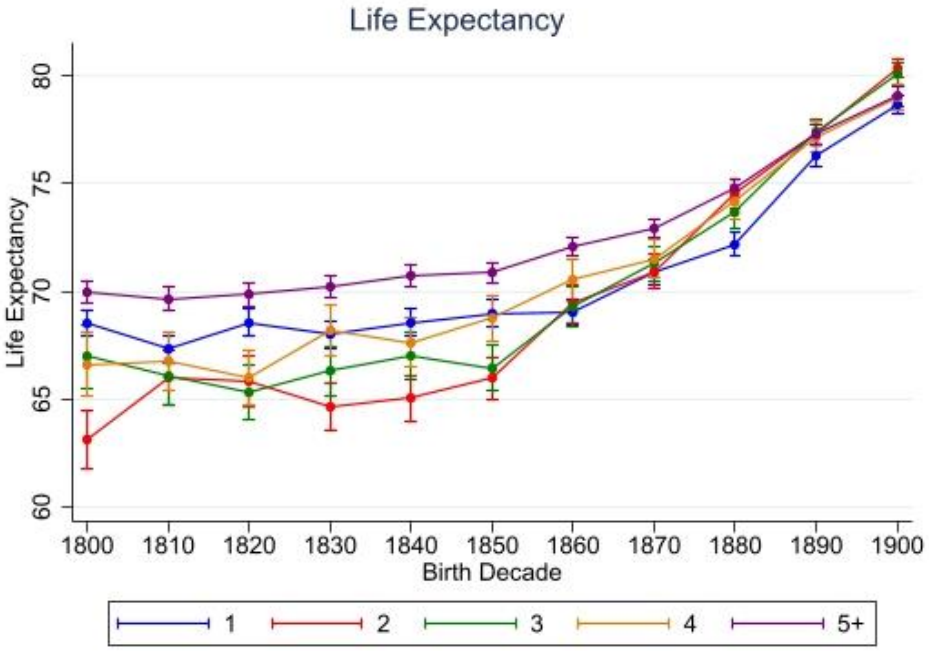

Figure 9. Predicted Life Expectancy at Age 50 by Number of Children, Males

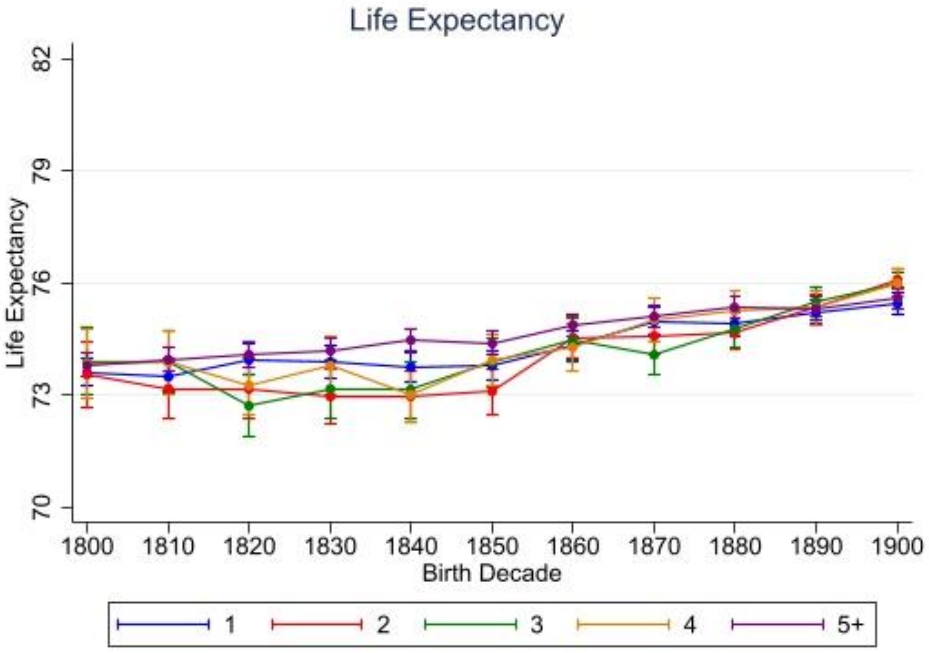

Figure 10. Predicted Life Expectancy at Age 50 by Number of Children, Females