Journal of Health Economics 67 (2019) 102224

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

XX > XY?: The changing female advantage in life expectancy

Claudia Goldin [a,*], Adriana Lleras-Muney [b]

[a] Harvard University, NBER, United States
[b] University of California at Los Angeles, NBER, United States

ARTICLE INFO

Article history:
Received 2 November 2018
Received in revised form 26 May 2019
Accepted 26 July 2019
Available online 10 August 2019

JEL classification:
J1
J16
N3

Keywords:
Life expectancy
Maternal mortality
Infectious disease

ABSTRACT

Females live a lot longer than males in most parts of the world today. But that was not always the case. We ask when and why the female advantage emerged. We show that reductions in maternal mortality and fertility are only partial reasons. Rather, the sharp reduction in infectious disease in the early twentieth century played a role. Those who survive most infectious diseases carry a health burden that affects organs and impacts general well-being. We use newly collected data from Massachusetts containing information on cause of death since 1887 to show that females between the ages of 5 and 25 were disproportionately affected by infectious diseases. Both males and females lived longer as the burden of infectious disease fell, but women were more greatly impacted. Our explanation does not tell us precisely why women live longer than men, but it does help understand the timing of their relative increase.

1. Introduction

Women live longer than men in most parts of the world today.[1] In many places, they live a lot longer. Among OECD nations in recent years, the difference in life expectancy at birth is around four to six years (seven in Japan). But have women always lived so much longer than men? The answer provided in recent studies is that they have not.[2]

Many hypotheses have been put forward to explain the so called "female advantage" in life expectancy. Most of the reasons why women live longer than men can be found by understanding the difference between having two X chromosomes, rather than having an X and a Y, and the resulting impact on hormones.[3] Whereas genetic factors are at play, the fact that the advantage increased greatly in the latter half of the twentieth century suggests that environmental factors, particularly those that interact with specific genetic elements, have disproportionately benefited women. In this paper, we examine *when* the female advantage emerged, to understand *what* caused the advantage to widen.

A complete answer concerning why there was a large increase in the female advantage still eludes us. But we argue that the reduction in infectious diseases played a role in extending longevity primarily through its impact on survivors and not because of the reduction in deaths directly from disease. To document this we collected data from Massachusetts containing information on causes of death from 1887 to around 1930 to show that people dying from infectious disease between the ages of 5 and 25 were disproportionately female. As infectious disease was reduced for all, females therefore reaped more benefits than did males. Though our explanation does not tell us why women were disproportionately affected by infec-

* Corresponding author.
E-mail address: cgoldin@harvard.edu (C. Goldin).

[1] See, for example, Barford et al. (2006), although the data are not entirely reliable for certain developing countries (and are dismal for some, particularly in Africa). On the possible reasons why women live longer, see Austad (2006) and Cullen et al. (2016).

[2] See, for example, Beltran-Sanchez et al. (2015), which uses birth cohorts for 13 nations in the Human Mortality Database. The female longevity advantage at age 40 and beyond appears consistently with cohorts born in the late nineteenth century. In France, for example, in the early nineteenth century women's life expectancy was only about one year greater than that of men.

[3] Mouse experiments have allowed reseachers to separate the impacts of chromosomes from hormones. See, for example, Arnold et al. (2012), Chen et al. (2012), and Du et al. (2014). Cardiovascular disease differs between males and females for both reasons. Males have more visceral fat whereas women have more subcutaneous fat, a difference determined by estrogen and also by the presence of the second X chromosome in females. Visceral fat stored in the abdomen predicts cardiovascular disease. Males, it has been shown in mouse models, are also more likely to suffer from hypertension even in the absence of different hormones. Sex differences exist in the incidence of autism-spectrum disorders, from which males are four to ten times more likely to be affected. Females are more likely to get auto-immune diseases, but their extra X chromosome protects them from more rapid degeneration once affected.

https://doi.org/10.1016/j.jhealeco.2019.102224
0167-6296/

tious disease, it helps us understand the timing of the onset of the female advantage.

We should be clear at the outset regarding the impact of infectious disease and its decline on the female life expectancy advantage. The direct effect will be small. It is the indirect effect among the survivors that we, and others, believe greatly explains the increase in longevity in general and the greatly widening advantage of females. As mortality from infectious disease fell, the fraction who were the survivors and who thus carried with them the markers of past illness, also decreased. A healthier population resulted, and females were disproportionately impacted.

Male mortality rates have always been higher for infants (Drevenstedt et al., 2008). But any advantage that females had at birth, or even those they had conditional on surviving to year one, were slight until some time in the twentieth century.[4] The female advantage in longevity greatly increased starting early in the twentieth century, reaching a maximum of eight years in the US by the 1970s. The difference then narrowed, but has remained substantial (Cullen et al., 2016).

The reasons that women began to live much longer than men have been studied by many. But existing explanations are incomplete. The reduction in maternal mortality as well as the decrease in the total fertility rate, for example, have been shown to explain at most one-seventh of the growing female advantage in longevity.[5] Our estimate is within that range. Increased smoking by men in the early twentieth century, which greatly expanded around WWII, is part of the reason that women began to live longer than men. But smoking was later taken up by women as well and that is one explanation for the narrowing of death rate differences by sex that has occurred since the 1970s.[6] Because smokers are affected with a 20- to 30-year delay, the full effect of the diffusion of tobacco use will not be apparent for some time. But even with this caveat, smoking alone cannot fully account for the rise and fall in longevity gender gaps (Cullen et al., 2016; Preston and Wang, 2006).

Our contribution is to explore the reasons behind the initial appearance and early widening of the female longevity advantage that started well before maternal mortality declined and prior to the mortality impact of the smoking upsurge. In the US, as well as in England (including Wales), France, and Sweden, a noticeable female advantage among those aged five to twenty years old emerged at the end of the nineteenth and beginning of the twentieth centuries. We connect our findings on this understudied phenomenon to the expansion of the female advantage that appeared later in the twentieth century.

Before the turn of the twentieth century, young women had a small but clear *disadvantage* in mortality from infectious diseases. That disadvantage disappeared when infectious disease prevalence fell, largely due to public health interventions. We examine various explanations for why girls had greater infectious disease mortality rates, but find no evidence that differential mortality was caused by differential nutrition by sex or by parental neglect. We argue that, because infectious disease in early childhood can be linked to chronic diseases in adulthood, the decline in infectious disease in the early part of the twentieth century generated a female advantage in childhood that later emerged as a female mortality advantage among adults in the second part of the twentieth century.

Most of the evidence we present is for the US and comes from our analysis of the vital statistics records for Massachusetts, the first US state to collect these data. These data are mainly in period, not cohort, form, but our work informs a cohort analysis. We conclude that the decrease in infectious disease as a cause of death for a cohort meant a decrease in the later-life burden of infectious disease on other causes of death for that cohort.

Our paper has four parts and proceeds as follows. In part I we review the long-run history of life expectancy for males and females in the US and Europe. Part II examines sex differences in mortality among the young and presents our findings about female excess mortality during periods of high infectious disease. Part III confronts why girls died at higher rates than did boys in the era of infectious disease. Part IV concludes with evidence on the relationship between the early disease environment and later-life mortality.

2. Life expectancy for males and females: the long run

2.1. The US: 1795 to 2015

Life expectancy at birth (e_0) increased greatly from the late nineteenth century—when it was around 50 years for females—to the mid-twentieth century—when it was around 70 for females. Much of that increase was due to the reduction in infant and child mortality. These facts are well known. Of greater interest, here, is that a female advantage appeared in the US sometime in the late nineteenth century.[7]

In each of the three life expectancy series for the US in Fig. 1, the female advantage emerged and expanded after the 1890s. The widening intensified in the first third to first half of the twentieth century, and particularly in the post WWII period, when growth in male life expectancy decelerated considerably. These trends reversed course in the 1970s when male life expectancy grew faster than that for females, and again in the 1990s when female life expectancy growth decelerated. The longevity gender gap at birth reached a maximum of eight years in the early 1970s and is about five years today.

The first of the Fig. 1 graphs (part A) gives life expectancy at birth. Until the 1890s life expectancy at birth was almost equal by sex. Part B, giving expectancy at age 15, shows a slight male advantage until the 1890s. The final graph (part C), giving life expectancy conditional on reaching age 45, shows a female advantage throughout the period, albeit initially quite small. (Note that in parts B and C longevity is expressed as the expected age at death, not as years left to live.) Based on this information it would appear that the female advantage at birth emerged in the US around 1890.

The shaded portion in each of the Fig. 1 graphs shows the period when the two series are connected: that from Hacker (2010) for the period to 1900, and that from the US Social Security Administration (SSA, 2005, 2017) for the years after. The Hacker series is always a bit higher than SSA's probably because the Hacker series is for whites only.

[4] Using the metric of life expectation at one year old or, say, at 15 years of age, females lived only about as long as males in many of today's rich nations prior to 1850, and even later in some. According to the Human Mortality database, life expectancy at age one was about equal for males and females in France in 1850 and only slightly higher for males in the UK. In Sweden, females appear to have lived longer than males ever since records have been kept.

[5] Retherford (1972), for example, showed that in the US the male-female gap at birth increased by 3.43 years in favor of girls between 1900 and 1960 and that decreased maternal mortality accounts for about half a year or 14 percent of the gap leaving a substantial unexplained portion. Our calculation, see Appendix Table A1, is around 15 percent. Albanesi and Olivetti (2016), citing Retherford's data, state that, between 1910 and 1965, all of the increase in the mortality difference between 20 to 39 years old men and women was due to maternal mortality decreases. But across all ages, the estimate is 14 percent.

[6] Beltran-Sanchez et al. (2015) show that smoking can explain about one-third of the difference between male and female mortality rates at ages 50 to 70 for cohorts born in the early part of the twentieth century. Others have found similar or larger contributions to smoking in explaining mortality gender gaps (Preston and Wang, 2006).

[7] We ignore, here, the far greater impact on male mortality of the American Civil War.

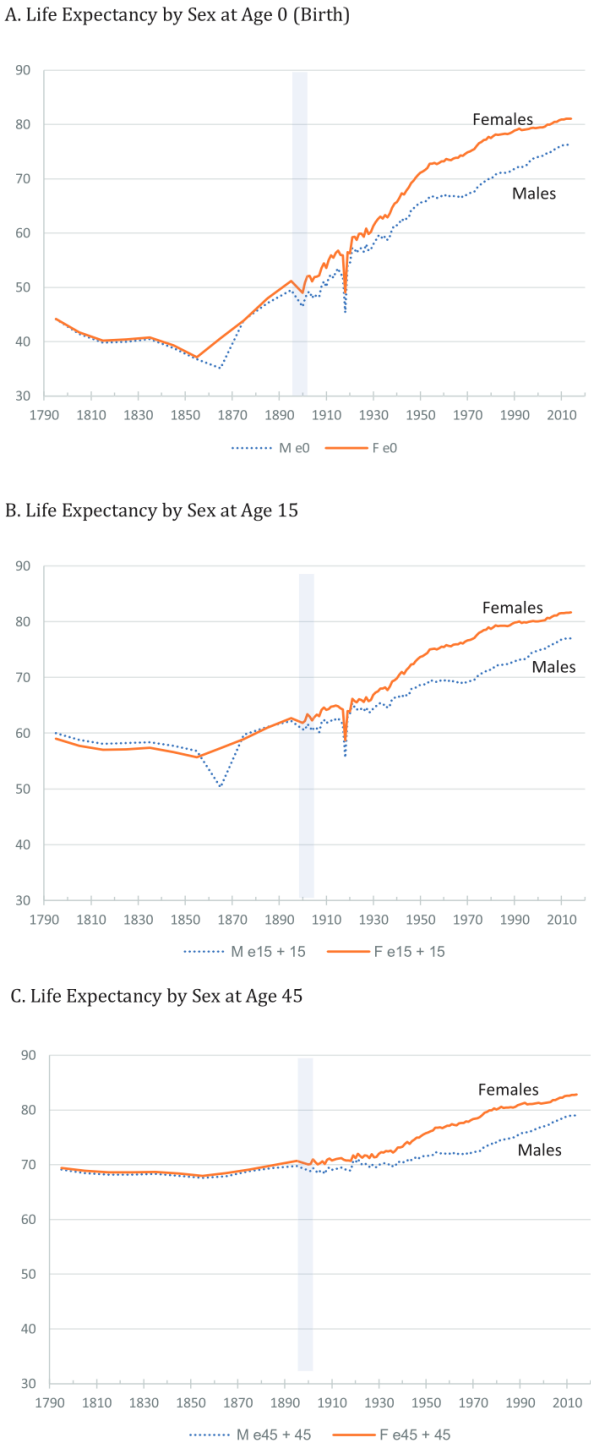


Fig. 1. Life Expectancy by Sex at Ages 0 (Birth), 15 and 45: United States.1795–2014.
A. Life Expectancy by Sex at Age 0 (Birth)
Notes: Shaded region gives the linkage points for the two series: Hacker for whites only and SSA for the entire population. No attempt has been made to link the series. The SSA data are used for 1900 and annual data after to 2015; the Hacker data are used for 1895 (1890–1900) and data on the quinquennial years prior. (e_{15} + 15), (e_{45} + 45) means that we add 15 or 45 years to obtain life expectation at birth rather than at the age given for consistency across the graphs.
Sources: Hacker (2010) for the white population until 1900; U.S. Social Security Administration (2005, 2017) for the entire population.

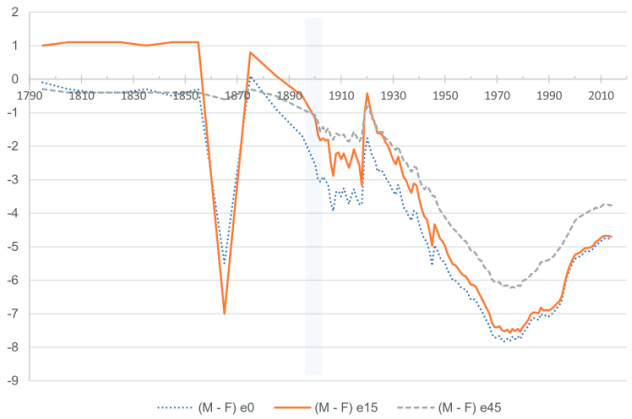


Fig. 2. Male Minus Female Life Expectancy: US.1795–2014.
Notes: Shaded region gives the linkage points for the two series: Hacker for whites only and SSA for the entire population. No attempt has been made to link the series. The SSA data are used for 1900 and annual data after to 2015; the Hacker data are used for 1895 (1890–1900) and data on the quinquennial year prior. The large relative decrease in male life expectation in 1865 is due to Civil War deaths and that just before 1920 is due to WWI casualties. The opposite impact around 1920 is due to the 1918 influenza pandemic. The main outlines of the graph do not change if it is expressed in relative terms (e.g., as a fraction of the male or female life expectation for each age).
Sources: See Fig. 1.

Trends in the female survival advantage can be seen better in Fig. 2, which gives the difference (male minus female) in life expectancy for the three ages, e_0, e_{15}, and e_{45}. Two large deviations from the story of an evolving female advantage can be seen. One is the US Civil War, which is responsible for the large sudden increase in male mortality around 1865. The other is the relative decrease in male longevity in the 1910s due to World War I. The opposite impact around 1920 is due to the 1918 influenza pandemic. Although we will show that the flu pandemic of 1918 produced a mortality disparity that relatively *disadvantaged* females, male war casualties overwhelmed it early on.

The conclusions from Figs. 1 and 2 are that the female advantage, conditional on surviving to age 15, appeared around the turn of the twentieth century. The female advantage at birth and conditional on surviving to age 45 were present even before. That said, the advantage widened considerably in all three cases after 1930 and then narrowed after 1970.

Males who survived to age 15 did better than females who survived to age 15 until the late nineteenth century. But females who survived to age 45 did a bit better than males who survived to age 45 throughout the nineteenth century. The mortality consequences of childbirth, either due to maternal mortality or the subsequent mortality consequences of maternal morbidity, may have caused the small differences between e_{15} and e_{45}.[8] But we will show that there are other reasons why women began to live a lot longer than men starting in the late nineteenth century. We also show that maternal mortality declines explain only a small fraction of the emerging female longevity advantage.

2.2. *Europe: early-nineteenth century to 2000s*

We provide similar life expectancy data and differences by sex at various ages for England (including Wales), France, and Sweden. These countries share some, but not all, of the features just described for the US. When measured by life expectancy at birth,

[8] On the morbidity consequences of childbirth, see Albanesi and Olivetti (2016).

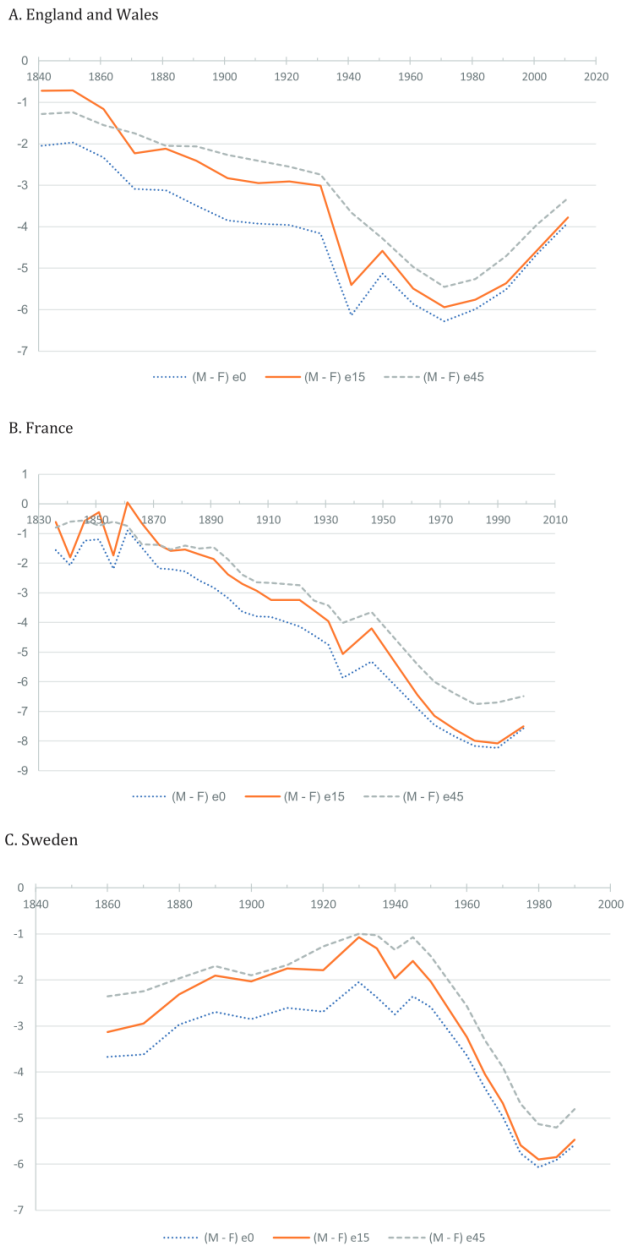


Fig. 3. Male Minus Female Life Expectancy (Period Rates): England and Wales, France, and Sweden, on Census Years.

Source: Human Mortality Database. Period tables. We only plot years for which censuses were conducted and thus population data by age are accurate. Census data are available every ten years starting in 1841 for England, and every ten years starting in 1860 for Sweden. For France they occured every 5 years between 1836 and 1936 (except for 1871 which was held in 1872, and for 1916 which was cancelled) and then for 1946, 1962, 1968, 1975, 1982, 1990 and 1999.

females in these nations had an advantage earlier than in the US. Recall that in the US, the female advantage emerged in the 1890s.

For the European data we use the Human Mortality Database (HMD) and present the data in Fig. 3 as the difference between male and female life expectancy at birth, and conditional on surviving to ages 15 and 45 (e_0, e_{15} and e_{45}), as we did in Fig. 2 for the US.[9]

Females in England (part A) and France (part B) had an edge on males from birth from the mid- to late-nineteenth century and that advantage then greatly expanded, as it did in the US.[10] Sweden (part C) is different. A female mortality advantage is apparent at all ages and as far back as the data allow us to go.[11] But these differences *shrink* to 1930. In all the countries considered, including the US, the female advantage greatly expanded after WWII and then started to decrease after the 1970s.

The reason for "Swedish exceptionalism" is not entirely clear but may stem from the low prevalence of infectious disease there in the nineteenth century. Swedish infant mortality rates were low through the nineteenth century and declined further beginning around 1810, far earlier than in England and France (Lynch and Greenhouse, 1994). Because infant mortality rates were largely due to infectious diseases, these patterns suggest that the female advantage emerged earlier in Sweden because the infectious disease environment was less severe.[12] But no data on mortality by cause of death exist to confirm the hypothesis.

Our discussion of the European mortality data has ignored the role of war largely because our figures include only years when census data were collected (and thus when we have accurate population counts), and censuses are generally not taken during conflicts. But large prolonged wars sharply increase death rates among young adults, significantly more so for men than women. Male deaths skyrocketed beginning in 1914 and returned to their previous levels only after 1919. In England, among 20 to 30 year olds, ten times more men than women died from 1914 to 1919. (Fifteen times more men than women in France in the same period).[13] Although women also died more during the war than at other times, the increase in their deaths was largely confined to the flu pandemic years.

3. Sex differences in mortality among children and young adults

3.1. Relative deaths by sex: Massachusetts

Historical evidence by age and cause of death for Massachusetts (MA) allow us to better understand aggregate US trends. MA has the longest and highest quality vital statistics data among all US states.[14] Death counts broken down by sex, age-group, and cause are available annually starting in 1885.[15] Identical data for the entire US population are available only from 1933, well after the female advantage appeared.

Fig. 4, parts A to D give the mortality series for males and females from 1887 to 1940 for four age groups: 5 < 10, 10 < 15, 15 < 20, and

[9] England and Wales, France, and Sweden are the most populous European nations with high quality vital statistics data from at least 1850. There are several other countries with early nineteeth century data in the HMD back to 1850, including Belgium, Denmark, Iceland, the Netherlands and Norway. But these countries are not as populous and, in consequence, the data are noisy. The HMD data are available online at http://www.mortality.org/. We will refer to England and Wales as England going forward.

[10] Cullen et al. (2016) report that male life expectancy exceeded female life expectancy at birth in 1900 in England and France also using the Human Mortality Database. The database has been updated and extended, and the data revisions account for the differences between our figures and theirs (personal communication).

[11] We plot only years with population censuses.

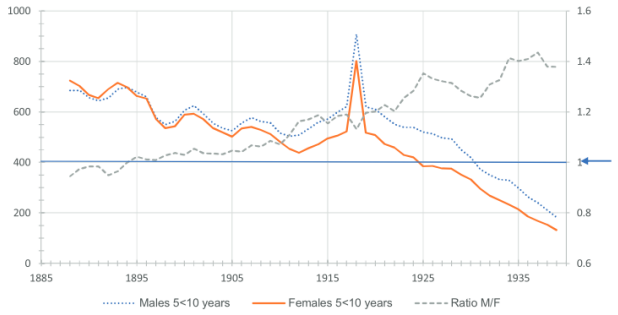
[12] An added question is why gaps in life expectancy shrank in Sweden before 1930 when they were expanding elsewhere. Cullen et al. (2016) suggest that gender gaps evolve with demographic and epidemiological transitions, and Sweden appears to have transitioned earliest.

[13] Results not shown but available upon request.

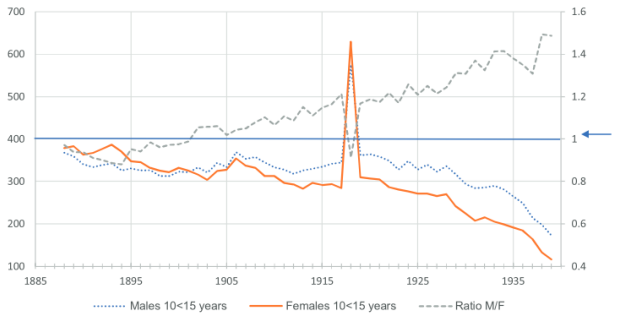
[14] Massachusetts, in 1842, was the first state to pass a state registration law.

[15] We should note that these are death records aggregated in various ways. Although the original data contain occupation, when the person had one just prior to death, the data we use do not have information on socio-economic status, work conditions, and other factors that might predict exposure to disease.

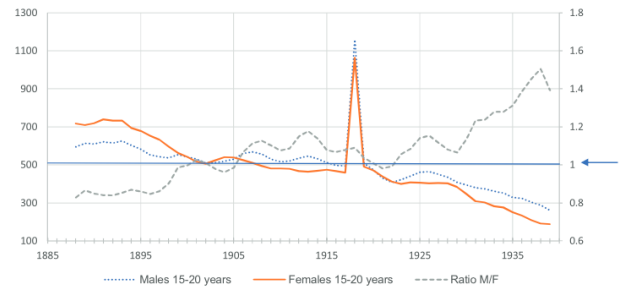
A. Male and Female Deaths, 5 < 10 Years Old and Male to Female Death Ratio (Right Axis)



B. Male and Female Deaths, 10 < 15 Years Old and Male to Female Death Ratio (Right Axis)



C. Male and Female Deaths, 15 < 20 Years Old and Male to Female Death Ratio (Right Axis)



D. Male and Female Deaths, 20 < 30 Years Old and Male to Female Death Ratio (Right Axis)

Fig. 4. Deaths by Sex and Age: Massachusetts, 1887 to 1940.
A. Male and Female Deaths, 5 < 10 Years Old and Male to Female Death Ratio (Right Axis)
Notes: Data are a three-year moving average of deaths for the age group except that 1918, the year of the influenza pandemic, is not part of any average. The two years before and after 1918 are averaged together.
Sources: Commonwealth of Massachusetts (1887 to 1940).

20 < 30 years.[16] Because of the relatively small number of annual deaths in some of these age brackets, the data are expressed as three-year centered moving averages. We exclude 1918 from the three-year averages, due to the large number of deaths from the pandemic, and include those deaths separately for that year.

There are three lines in each graph. Two give the number of male and female deaths in each year for the relevant age group. The dashed line gives the ratio (Males/Females) of the two series and is mapped on to the right axis. Equality of the ratio is given by a thicker line at one and an arrow.

Female deaths exceeded male deaths to around 1900 for all four age groups.[17] Also important is the sharp decrease in the male to female death ratio in 1918 for 10 < 15 year olds as well as a smaller decrease for the 5 < 10 year olds. Females between 10 and 15 years old were dying at higher rates than males in 1918. It is possible that these differences extend to the 15 < 20 year group, but deaths of young men in World War I are a confounding factor and could mask the effect of the flu. Interestingly, the gender death ratio was less than one for 20 < 30 year olds from 1918 to 1935.[18]

The main conclusion is that female children, especially those between 10 and 15 years old, died at higher rates in 1918 from the flu than did males. Relative deaths after the flu pandemic resumed the previous trend of an increasing number of male than female deaths, thus widening the female advantage.

The fact that more girls 10 < 15 died from the virulent flu of 1918, suggests that young females may have also died at higher rates from other infectious diseases, at least after age one.[19] That is what we found. Public health interventions that reduced the burden of infectious disease, therefore, disproportionately decreased female (non-infant) deaths and increased female life expectancy relative to male life expectancy.[20]

Of equal interest is that the 20 < 30 year old group in MA shows a distinct break around 1920. Previously, women had died at lower rates than men in those ages. But after 1920, both men and women in that age group died at significantly lower rates. The reason for the decrease for both sexes, we will soon demonstrate, is that deaths from infectious diseases of all types, but particularly from water-borne diseases, decreased substantially around 1900. Deaths in childbirth, however, had not yet decreased and therefore the remaining deaths in the 20 < 30 year old group tilted relatively toward women.

[16] We begin the series with 1887 because the probable cause of death was aggregated in some meaningful manner beginning in that year.
[17] These are the absolute number of deaths in all of Massachusetts, by gender, in each year.
[18] Noymer and Garenne (2000) attribute the temporary decline in the female advantage among adults to the delayed effects of the influenza pandemic. The pandemic, according to them, may have killed weaker males thus lowering mortality rates in subsequent years among male survivors.
[19] Women younger than 25 years succumbed to pulmonary tuberculosis at higher rates than did men of the same ages in the early twentieth century (Hacker, 2010; Henry, 1989; Madigan, 1957). The opposite is true among those older than 25 years—males appear to have been more susceptible to TB in the early twentieth century (Noymer and Garenne, 2000, Fig. 7). Note that the greater infant mortality rate for males than females is attributed to a greater susceptibility to infectious diseases among male infants (see, e.g., Drevenstedt et al., 2008). But after around age five, females appear to be more susceptible to infectious diseases.
[20] On public health measures that began in the late nineteenth century and gained greater importance in the early twentieth century, see Cutler and Miller (2005) on the role of filtration and chlorination of water, and Alsan and Goldin (2018) on water and sewerage systems. Omran (2005) discusses the notion that the overall epidemiologic transition from infectious to chronic disease favored women. Beltran-Sanchez et al. (2015) note that females may have some inherent vulnerability to infectious disease relative to males after infancy, but that they have various genetic protections against chronic diseases, except those in the autoimmune category.

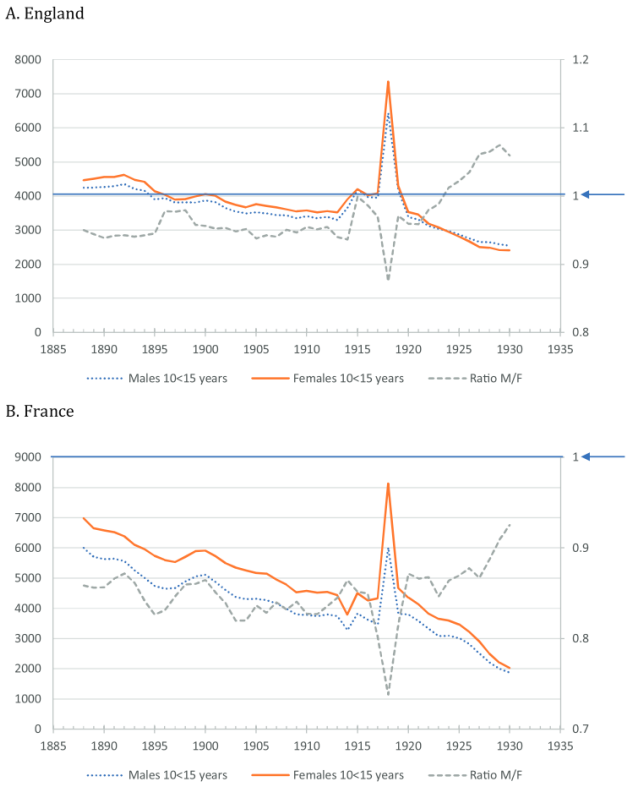


Fig. 5. Male and Female Deaths, 10 < 15 Years Old and Male to Female Death Ratio (Right Axis): England and France, 1880s to 1930.
Source: Human Mortality Database. Period tables.

3.2. Relative deaths by sex: England and France

The female advantage among youth 10 < 15 years old appears later in England and France than in MA. Recall that the female advantage in life expectancy in England and France occurred earlier than in the US. But, as can be seen in Fig. 5, deaths among male youths 10 < 15 in England started to exceed deaths of female youths 10 < 15 only by 1925. In France, male deaths in that age group never exceeded that for females.

Fig. 5 has been drawn in a manner similar to that of Fig. 4 for MA.[21] In a parallel manner to the MA case, girls suffered greater losses than boys during the flu pandemic years in England and France, resulting in a sharp decline in the ratio of male to female deaths. As in the MA case, by looking at the evolution of mortality by sex for 10 < 15 year olds, we can remove some of the direct effect of the war because those males would have been too young to serve. Since the pattern is similar to that for MA, it is probably due to a greater susceptibility of young females to the flu, though we cannot fully separate the effect of the war in Europe from that of the flu.

3.3. Death rates by sex: MA and Europe

The main finding from Fig. 4 is that deaths of young males and females in the US became equal around 1895 for the age groups given. These data are suggestive of an emerging female advantage. But the evidence is not definitive because the data have not been

[21] To mirror what we do for MA, we construct three-year moving averages, except for the war years, which are not part of any average. Sweden is omitted because of its much smaller population.

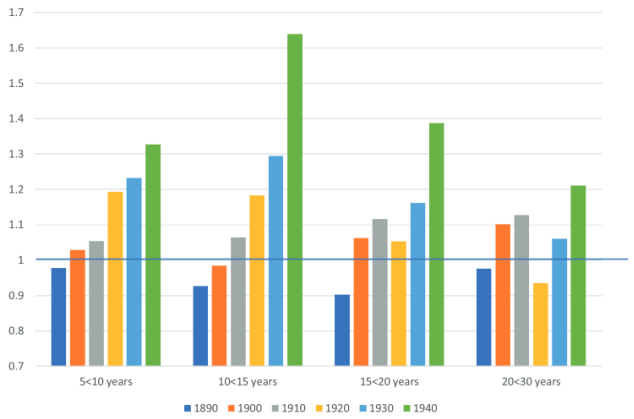


Fig. 6. Ratio of Male to Female Death Rates for Various Age Groups (5 < 10–20 < 30): Massachusetts.1890–1940.
Notes: Three-year centered moving averages for the death rates are used for data from 1887 to 1941. Relative death rates are given by: (Male deaths in interval/Male population in interval)/ (Female deaths in interval/Female population in interval)
Sources: Commonwealth of Massachusetts (1887 to 1940). U.S. Department of the Interior, Census Office, 1890 to 1900; U.S. Department of Commerce, Bureau of the Census, 1910 to 1940.

expressed as a fraction of the population. In fact, the population of males and females by age deviated from equality in MA in the 1920s. Textile production in the US shifted from the northeast to the south and the manufacturing economy of the Commonwealth declined. In consequence, young men left the state in greater numbers than did women.[22]

Source: U.S. Department of Commerce, Bureau of the Census, 1910 to 1940.

We now use population data at census intervals to create death rates for the four age groups 5 < 10, 10 < 15, 15 < 20, and 20 < 30 years.[23] Accurate population estimates between census years can be computed only with additional information on migration, which does not exist at high enough frequency by age and sex.

Fig. 6 provides the death rates for males relative to females. The relative rates show that the female advantage did not exist in 1890 for any of the four age groups, but that it did exist for three of them by 1900 and for all by 1910. In the case of the three younger groups shown, the trajectory from 1890 to 1940 shows that the relative gains made by females continued to 1940 at least.

For the two age groups 15 < 20 and 20 < 30, the flu pandemic and WWI break the trend. Furthermore, by 1930 the burden of infectious disease had become sufficiently low, as noted before, that only persistent maternal mortality prevented the female death rate from continuing its relative descent. But after 1937, with the introduction of sulfa drugs and soon after with the innovation and diffusion of antibiotics, the female death rate in the prime reproductive ages became substantially less than that for men.[24]

Fig. 7 (drawn to the same scale as Fig. 6) shows similar mortality rate data for England from roughly 1880 to 1930 taken from Davenport (2019). The female advantage for these four ages groups

[22] The ratio of males to females in the 20 to 30 year groups for Massachusetts declined in the 1910s and 1920s and then recovered in the 1930s.

Age Group	1910	1920	1930	1940
15<20	0.97	0.96	0.98	1.00
20<25	0.93	0.89	0.89	0.95
25<30	0.97	0.94	0.90	0.94

[23] Death rates are estimated at decade intervals because of the need for accurate population data.

[24] On the role of sulfa drugs in reducing maternal deaths in the US, see Jayachandran et al. (2010).

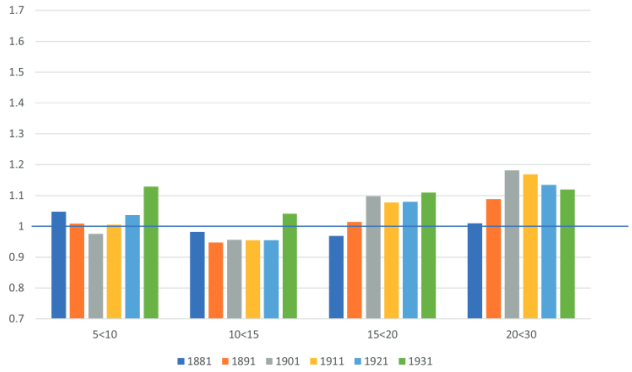


Fig. 7. Ratio of Male to Female Death Rates for Various Age Groups (5 < 10–20 < 30), England:1881–1931.
Source: Annual deaths by cause, Age and Sex in Englad and Wales, 1848–1900. UK Data Archive Davenport (2019).

also appears sometime around the turn of the century (except for the 10 < 15 group). But the advantage in MA, when it appeared, was considerably greater for all ages. By 1930 and 1940, the ratio of male to female mortality below age 20 in MA was about 1.2 or higher, but the ratio never exceeded 1.2 in England.

3.4. Cause of death by sex

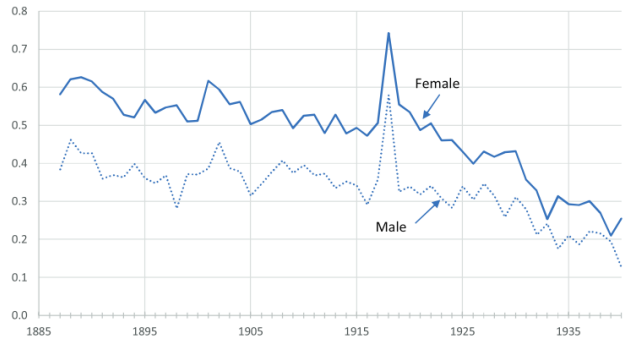
The female longevity advantage, as we have shown, appeared in the US and in MA around the 1890s. We now demonstrate that infectious disease as a cause of death declined significantly after 1890, as public health measures, such as clean water and sewerage systems, spread across municipalities and states (Alsan and Goldin, 2018).

Various problems arise in categorizing cause of death in the past. The most important is that cause of death was not always known (occasionally not even today) and diseases can be manifested in a variety of ways. As knowledge advanced, cause of death categories became more accurate. In the MA data, cause of death categories increased from six major groups starting in 1850, to 14 in 1901, 15 in 1921 and 18 in 1931.

A far greater fraction of deaths of females than of males in the 10 < 15 year group was due to infectious disease throughout the period, as can be seen in Fig. 8, part A. But more males than females died from external and violent causes. Except for the years after 1925 and for those with significant infectious disease epidemics, the fraction of non-violent deaths, rather than all deaths, from infectious disease was higher for females than for males, as can be seen in Fig. 8, part B. Another important factor is the fraction of all deaths (including or excluding those caused by violent factors) from infectious disease greatly decreased with time and decreased more for females than males.[25]

The data in Fig. 8, like those in Fig. 4, are not expressed as rates. To demonstrate the decreasing importance of infectious disease as a cause of death, we divide by the appropriate population. In Fig. 9 we express deaths as a fraction of the age and sex of the relevant population groups for census years, and use only deaths from non-violent factors for the 10 < 15 year group.[26] A clear decrease in the death rate can be observed for both males and females during the

A. Fraction of All Deaths due to Infectious Diseases



B. Fraction of Non-violent Deaths due to Infectious Diseases

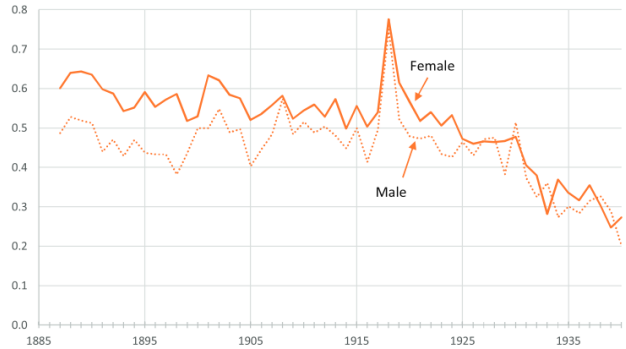


Fig. 8. Fraction of Deaths of Males and Females 10 < 15 Years Old Due to Infectious Diseases: Massachusetts, 1887 to 1940.
A. Fraction of All Deaths due to Infectious Diseases.
Notes: The lines in Part A are (infectious diseases/all deaths); the lines in Part B are (infectious diseases/all non-violent deaths). Infectious diseases are defined as: 1887-1900: Zymotic and Constitutional; 1901-1920: General and Respiratory; 1921-1930: Infectious, General and Respiratory; 1931-1940: Infectious and Parasitic; Respiratory. Actual population values are used rather than three-year moving averages.
Sources: Commonwealth of Massachusetts (1887 to 1940).

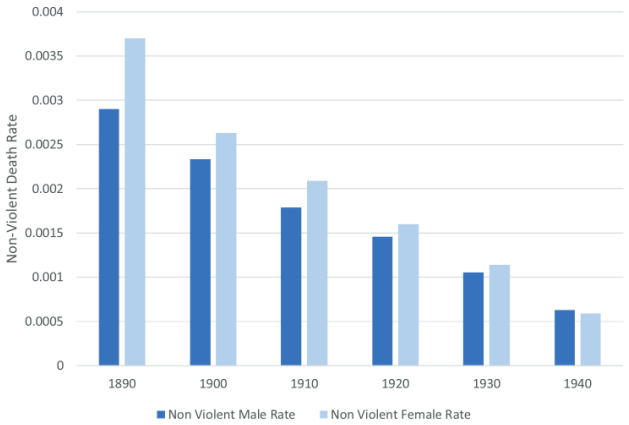


Fig. 9. Death Rate from Non-Violent Deaths: 10 < 15 Years Old: Massachusetts 1890 to 1930, by Sex.
Notes: Three-year centered moving averages are used for non-violent deaths where the centers are chosen as 1890, 1902, 1910, 1922, and 1930 to aggregate within cause of death aggregates.
Sources: Commonwealth of Massachusetts (1887 to 1940); U.S. Department of the Interior, Census Office, 1890 to 1900; U.S. Department of Commerce, Bureau of the Census, 1910 to 1940.

[25] Johansson (1984) notes that tuberculosis hit young females harder than it did young males. Smith (2008), in her in-depth study of four western Massachusetts rural towns, finds that females 10 to 19 and 30 to 39 years old had a clear mortality disadvantage with respect to TB until around 1885.
[26] Given the imprecise nature of cause of death, we chose to use all deaths due to non-violent causes. For ages 10<15 the majority were due to infection, although some may have been described as due to diseases of certain organs.

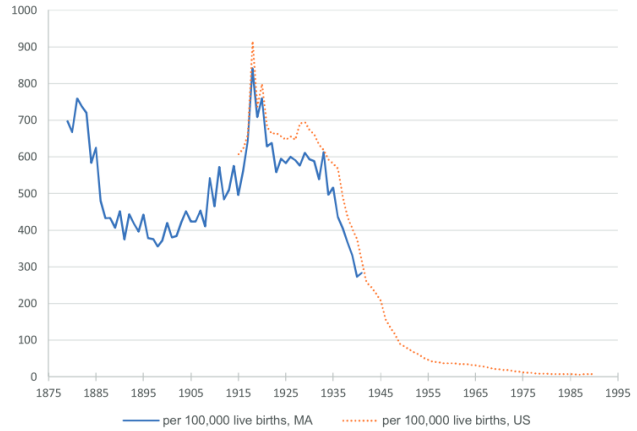


Fig. 10. Maternal Mortality Rate in Massachusetts: Maternal Deaths/100,000 Live Births, 1887 to 1941.
Notes: For cause of death categories, see notes to Fig. 11. This figure almost perfectly matches one in Loudon (1992, p. 381, fig. 22.6), although Loudon does not give the underlying series. The one year when the series do not match (1890) stems from a copying error by Loudon apparently due to poor original print quality.
Sources: Commonwealth of Massachusetts (1887 to 1940); *Historical Statistics* (2006), series Ab924.

1890 to 1940 period, although the decrease is larger for females. Because infectious disease was a more important cause of female than male deaths, females were disproportionately advantaged as it declined.

3.5. Role of decreased maternal mortality in the female advantage

We noted before that infectious disease decreased for much of the period but that maternal mortality in the US and MA did not decrease greatly until the mid-1930s with the advent of sulfa drugs. The maternal mortality series for the US begins in 1915, but the MA series can start in the 1870s because the Commonwealth's vital statistics provide sufficient detail on cause of death.

Our maternal mortality series for MA, given in Fig. 10, is defined as the number of deaths from maternal causes per 100,000 live births.[27] We also include the standard US series from its start in 1915 to 1990. In late nineteenth century MA there were 400 maternal deaths per 100,000 births. But the rate rose in the early twentieth century to a peak during the flu pandemic period, after which it declined beginning in the early 1930s. The MA series is generally lower than that of the US, but it is still far higher than those of England and Sweden in 1930.[28]

As infectious disease began to wane, the general death rate decreased in the 1890s. But maternal mortality did not decrease until much later and rose for some of the period. Thus, the fraction of female deaths due to maternal causes increased, as shown in Fig. 11. The fraction of all deaths of women 20 < 30 years old from maternal causes was less than 6 percent in 1888, but was almost 18 percent—three times more—in 1930. (See the dotted line graphed to the right axis.) After the introduction of sulfa drugs, deaths due to maternal causes fell, and the fraction due to those causes began its descent, as can be seen at the end of the series.

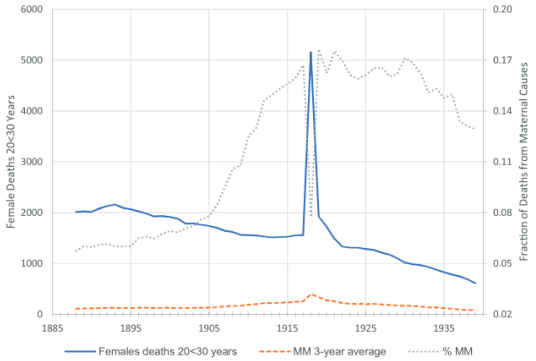


Fig. 11. Female Deaths, 20 < 30 Years Old, Deaths due to Maternal Causes, and the Fraction of All Deaths due to Maternal Causes: Massachusetts, 1887 to 1940.
Notes: "Female deaths 20 < 30 years" is a three-year centered moving average of all deaths, with the exception of 1918. The two years around 1918 are expressed as a two-year moving average and 1918 is not averaged. "MM 3-year average" is the number of deaths to women 20 < 30 years old attributed to childbirth expressed as a three-year moving average. The year 1918 is excluded from the three-year average. Cause of death aggregates changed in the Massachusetts vital statistics three times in the period examined: 1901, 1921, and 1931. "% MM" is the fraction of all deaths attributed to maternal causes.
Cause of death subcategories attributed to maternal causes:
1887-1900: Abortion, childbirth, miscarriage, puerperal convulsions.
1901-1920: Accidents of pregnancy; hemorrhage, puerperal; other accidents of labor; septicemia, puerperal; albuminuria and puerperal eclampsia; phlegmasia alba dolens, puerperal; other puerperal accidents, sudden death.
1921-1930: Accidents of pregnancy (e.g., abortion, ectopic gestation); puerperal hemorrhage; other accidents of labor (e.g., Caesarean section, other surgical operations and instrumental delivery); puerperal septicemia; puerperal phlegmasia alba dolens, embolus, sudden death; puerperal albuminuria and convulsions following childbirth; puerperal diseases of the breast.
1931-1940: Abortion; ectopic gestation; puerperal hemorrhage; puerperal septicemia and pyemia; puerperal tetanus; puerperal albuminuria and eclampsia; other toxemias of pregnancy; puerperal phlegmasia alba dolens, embolus, sudden death; other accidents of childbirth (e.g., Caesarean operation); other and unspecified conditions of the puerperal state.
Sources: Commonwealth of Massachusetts (1887 to 1940).

Given the large decrease in maternal mortality across the twentieth century, as well as the decrease in the number of births per woman, one might have thought that the decrease in each separately or together was responsible for a major part of the increasing female advantage, particularly after the 1930s. Even though maternal mortality fell sharply, only about one-seventh of the rising female advantage was due to the decrease in maternal mortality. Even for this fertile age group, TB, not maternal mortality, was the leading cause of death for females in 1930 (Enterline, 1961).[29]

The finding that just one-seventh of the change can be explained by decreased maternal mortality is made clear in the following hypothetical exercise. If the maternal mortality rate in 1900 were decreased to the rate that existed in 1990—almost zero—life expectancy for women in the US would only have increased by about 0.6 years at birth and 0.7 years at age 20. (See Appendix Table A1.) The gain from eliminating maternal mortality pales in comparison with the actual increase in life expectancy from 1900 to 1990 (see Fig. 1): 29 years at birth (50 to 79 years), 18 years at age 15 (62 to 80), and 11 years at age 45 (70 to 81).[30]

[27] See Fig. 11 for the list of deaths due to maternal causes. Our time series is identical to that in Loudon (1992, p. 381, fig. 22.6). Loudon did not provide the underlying series for the numerator and denominator separately, but because we use the same vital statistics series as he did for births and deaths by cause, we can replicate his series.

[28] Loudon (1992) attributes the lower maternal mortality in Europe to their greater use of midwives than in the US, as well as to fewer interventionist practices in childbirth.

[29] See also Albanesi and Olivetti (2016, p. 657). Moreover TB killed more females than males in the US for those 5 to 25 years old (Doege, 1965). Johnston (1995) shows that female mortality rates from TB exceeded male mortality rates from TB during 1900 to1930 when TB rates were high. According to Johnston (1995): "In populations where TB is on the rise more females than males die of it, although the reasons for this remain unknown. Conversely, where the disease is on the decline male mortality from tuberculosis is usually greater than female."

[30] Preston (1976) does a similar calculation (see his table 3.3 and related discussion).

More to the point of explaining the growing female advantage, the gap in life expectancy between males and females at birth was about three years in 1900 but was seven years in 1990. Thus the decrease in maternal mortality of around 0.6 years explains 15 percent of the increased female advantage at birth. That is certainly an overestimate because the medical advances that greatly reduced maternal mortality—antibiotics, blood transfusions, antiseptic operating, and advanced surgical techniques—would also have increased male life expectancy. And they would have improved the life expectancy for females in general.

4. Why girls died at higher rates from infectious disease before c.1890 in the US

An important finding from our work is that that life expectancy at birth was not more favorable for females in the US until the 1890s. Aside from maternal mortality, a primary reason is that female children died from infectious disease at higher rates than did male children. One possibility is that daughters were less well fed than were sons, and may have disproportionately succumbed to infectious disease in the pre-public health era. If that was the case, the changing mortality position of females would have been due to rising income levels that freed the constraints on families. But it appears from our analyses, as well as those of others, that this was not the case.

Two other possibilities remain. One is that female children had a greater role taking care of sick family members whereas the boys were out of the house more, possibly at work. Another is that female children had a greater inherent susceptibility to infectious disease and, possibly, were less able to fight the worst of the infectious maladies. In both cases, the difference between males and females would have greatly declined as the environmental burden of infectious disease fell overall.

We should note that others have also pointed to the greater mortality of young females in the period when infectious disease was a major killer. The distinguished demographer George Stolnitz (1956), for example, noted that prior to the 1920s females died at higher rates than males extending from childhood to mid-life. Stolnitz correctly emphasized several misconceptions about gender differences in mortality. One was that females always outlived males. Another was that the only period in which females did not have an advantage was during their childbearing ages.[31] Both were incorrect.

Pinnelli and Mancini (1997) produce similar findings to ours for Italy, except that excess female mortality in Italy did not disappear in the relevant age group until the 1920s to 1930s. They do not definitively say why there was excess mortality among young females, and downplay relative deprivation. They attribute excess female mortality to the fact that girls were more likely to be at home with sick family members.

The evidence to negate the hypothesis that female children suffered from relative deprivation comes from several sources and methodologies. One is sex ratios across the US and the sex of the last child. Another is expenditures on female versus male children. Yet another is inferred from anthropometric data on heights and weights, in our case from a comparison of native-born and Irish parentage school children in 1872 Boston. The evidence suggests

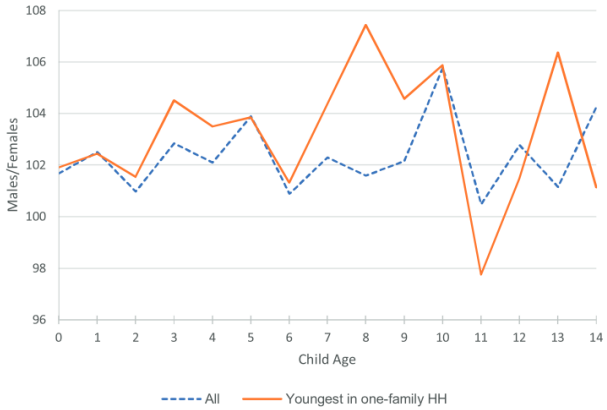


Fig. 12. Child Sex Ratios (Males/Females) in US Families:1850–1900.
Notes: To determine the youngest in the family we restrict the sample to individuals living in one-family households who are the biological children of the household head (we drop minors who are adopted, spouses, and those having another relationship to the household head). The line labeled "all" includes all children in all types of families.
Source: Ruggles, et al. (2019). 1850, 1860, 1870, 1880 and 1900 Censuses, 1 percent random samples.

that US girls in the nineteenth century were not relatively disadvantaged nutritionally or otherwise.

The literature on sex ratios has some detractors, but it appears that support for male (or son) preference in the nineteeth century US—that parents wanted to have boys rather than girls—is weak. Hammel et al. (1983) claim that differences in sex ratios by place show that girls were more valued where their wages were higher. Courtwright (1990) argues that the slight imbalance of sex ratios (western states having more male children than eastern states) was mainly due to selective migration and that there is no evidence to support the contention that it is indicative of poorer treatment of girls in places that did not value their production as much as in more industrial areas.

We use the 1850 to 1900 US population censuses to assess whether parents had son preferences. We follow the intuition from Barcellos et al. (2014) and investigate the sex ratio by age *among the youngest* in the family. If parents continue to have children until they have a boy (or a certain number of boys), we will observe that the youngest child in the family will tend to be a boy and the sex ratio among the youngest will rise with age. The test is imperfect because the sex ratio would also rise with age if girls died at higher rates for biological reasons. But we can assess that directly by comparing the sex ratio by age in the overall population with the sex ratio among the youngest in the family.

Fig. 12 shows sex ratios (the number of males divided by the number of females) by age for different subsets of children ages 0–14 years. A sex ratio at birth of around 105 males per 100 females is considered normal by demographers. Although the data are noisy, the sex ratio does not systematically increase with age and, most importantly, does not rise among the youngest.[32] The highest sex ratio we observe is 107. By comparison, the sex ratio in India among the youngest is 138 by age four (Barcellos et al., 2014).

Another test of son preference is given by sex-ratios at birth, by birth order of the child. If parents prefer sons, infanticide and

[31] According to Stolnitz (1956, p. 24): "Among nineteenth-century Western populations, for example, the highest frequency [of lower male mortality rates]—well over 50%—is encountered in the pre-reproductive age interval of about 7 to 12; the next highest percentage is for 12 to 17, when fertility is very much lower than at subsequent ages."

[32] Data on fertility histories for every woman would be required to properly conduct the test, but that information is not available in the census. Instead we focus on biological children of household heads who live in one-family households, and infer birth order using their ages. We then plot the sex ratio among the youngest. We restrict attention to children under 14 years living in household quarters, who are listed as a "child" of the household head. Then we order children by age and identify the youngest in the family.

Table 1
Sex Ratios in the US:1850–1900.

	Observed Birth Order of child				
	1	2	3	4+	N
Part A: Sex ratios among Census population in US 1850-1900					
Ages 0–20	1.065	1.041	1.037	1.023	751,403
Less than age 2	1.006	1.019	1.043	1.023	86,562
Part B: US 1850 census only					
Ages 0–20	1.053	1.054	1.044	1.051	73,407
Less than age 2	0.946	1.038	1.065	1.009	8,070

Sources: Ruggles et al. (2019). IPUMS, 1850, 1860, 1870, 1880 and 1900 Censuses, 1 percent samples.
Notes: Unweighted ratios are reported. To compute birth order, we restrict attention to individuals living in one-family households who are the biological children of the household head (we drop minors who are adopted, spouses or have another relationship to the household head). We then assign birth order based on age of the existing children in the household. Birth order is therefore likely to be measured with error because of child deaths and also because older children may not have been living in the household—this should make the pattern across ages stronger since gender preferences become more pronounced with higher order births.

abandonment of infant girls should be more common at higher-order births. We do not observe sex by birth order at the time of birth, but we can infer birth order within households using child age. The inference would be incorrect if children died or moved away, but it provides a useful approximation. Table 1 shows sex ratios by observed birth order among all children to 20 years old, and also among infants, for the combined population censuses of 1850–1900 and for 1850 alone. In the nineteenth century US, sex ratios did not greatly or systematically increase with birth order.[33]

We also investigate other indicators of son preference and find no evidence for it in the US data.[34] In countries with son preference, girls are abandoned at a greater rate than boys. In the US we see no statistically significant difference in the fraction of males and females living in households in 1850, and when we pool censuses from 1850 to 1900, we find that males were more likely to be living in group quarters, such as institutions for the poor, and less likely to be living in households compared with girls. (See Appendix Table B2.)

Another indirect test of son preference is whether girls live in larger, poorer families compared with boys. As argued by Jensen (2002), if families continue to have children until they reach their desired number of boys, then on average girls will have more siblings, and these families will tend to have fewer resources per child. We do not find confirming evidence of that effect and find the opposite result for 1850. (See Appendix Table B2.)

Even in the absence of son preference, parents could treat boys and girls differently, and potentially discriminate against girls, causing them to be less healthy and increasing their death rates. The historical evidence on differences in family expenditures on sons and daughters is weaker, and we have found none on food allocations.[35] Logan (2007) uses the 1900/01 consumer expenditure study to show that there does not appear to have been gender bias in the allocation of parental expenditures to children by gender. In addition, there are no statistically significant differences in the rate at which boys and girls (ages 5 to 14) attended school in 1850 and 1880. In the early part of the twentieth century, in fact, girls attended and graduated from high schools at a far greater rate than did boys.[36]

To evaluate the proposition that female children experienced relative nutritional deprivation in poor households, we use anthropometric evidence from Bowditch (1877), a study of 24,500 Boston public school children in 1872. The children were weighed and measured. Height was taken without shoes and allowance was made for the weight of clothing. Nationality and race were recorded. Although there were too few blacks to be reported separately, there were a large number of Irish. The study was motivated by interest in the growth of girls relative to boys at the time of puberty, to discover reasons for "the alleged inferiority in physique of American women" (p. 4).

We have chosen to use the white native-born children with native-born parents as the "control," or standard, and the (mainly native-born) children of Irish-born parents as the "treatment." The Irish, being poorer than the white native-born, would have faced greater resource constraints. Because they also had more children, they would have had greater difficulty providing for their families. But did their lower incomes lead to relative deprivation for their daughters? Were Irish girls shorter and thinner relative to Irish boys in comparison with native-born girls relative to native-born boys?

To find the answer to that difference-in-difference question, we run the regression given by eq. (1), where i = sex and j = nativity and k = age. We assess whether the coefficent δ on (female × Irish) is significant and of substantial enough magnitude. The outcomes, y, are either height or weight. Each is expressed three ways: absolutely, in logs, and as Z-scores by age for each sex.[37] Since the Irish sample is small at older ages—probably because many older Irish children did not attend school—only children five to ten years old can be analyzed. Age dummies, θ, are included. We provide the results in Table 2.

$$y_{ijk} = \alpha + \beta\, Female_i + \gamma\, Irish_j + \delta\left(Female_i \times Irish_j\right) + \theta_k + \varepsilon_{ijk} \quad (1)$$

We should note that all the children in the Bowditch sample were short and thin by modern standards (the means of the Z-scores are negative, indicating the children were below the modern CDC standard). For each of the three height and weight measures the Irish are shorter and lighter—consistent with our sense that they were less well nourished than native-born children. But the girls of Irish parentage were not much shorter or lighter relative to native-born girls. Although the coefficients on the interaction terms are negative, the standard errors are large and the magnitudes are small. Girls, in the entire sample, were not significantly shorter than males relative to modern standards (see col. 3, Z-score), although they were somewhat lighter in weight.

We find, therefore, no compelling evidence of relative deprivation of girls among poorer households, using the Irish as the treatment and the native-born parentage group as the control. And

[33] We also estimated regressions to assess whether the family's fertility depended on the sex composition of their existing children. In India and China, if the first born is female, families are far more likely to have a second birth. If the two first-born are female, parents are again more likely to continue to have children. In the US in the 1850s, we find the opposite pattern. Families that have boys are more likely (rather than less likely) to have more children. Zeng et al. (1993) shows that in modern China, sex ratios rise to 131 for the fourth-born child.

[34] We find that fathers are more likely to remain if the first born is a daughter, which is inconsistent with the more recent findings in Dahl and Moretti (2008) that mainly concern shot-gun marriages.

[35] The development literature has considerably more direct evidence on son preference (e.g., Jayachandran and Pande, 2017). On adult female mortality and deprivation historically, see Klasen (1998) on Germany, which uses indirect evidence on remarriage and value of women's work.

[36] On the schooling of 5 to 14 years olds in 1850 and 1880, see Goldin and Katz (2008, Fig. 4.2, 153). On attendance and graduation from high school, see Goldin and Katz (2008, Fig. 6.5, 231). Appendix Table B2 also presents results for the 1850 Censuses alone and for 1850-1900, showing no economically or statistically significant differences by gender in school attendance.

[37] We use the LMS method as implemented by the zanthro command in Stata to estimate Z-scores relative to the modern standard, as given by the 2000 CDC growth charts. The LMS method has become the standard for use with anthropometric data because usual Z-scores assume normality. The CDC modern height and weight standards can be found at: https://www.cdc.gov/growthcharts/percentile_data_files.htm

Table 2
Heights and Weights of School Children, 5–10 years old: Boston, 1872.

	Heights (in inches)			Weights (in pounds)		
	(1) Height	(2) Log(Height)	(3) Z-Score	(4) Weight	(5) Log(Weight)	(6) Z-Score
DependentVariable Mean	46.78	3.84	−1.196	50.81	3.91	−0.681
Female	−0.278***	−0.00595***	0.0425	−1.780***	−0.0362***	−0.180***
	(0.0692)	(0.00148)	(0.0318)	(0.212)	(0.00399)	(0.0280)
Irish	−0.460***	−0.00958***	−0.207***	−0.568***	−0.00862**	−0.0506**
	(0.0599)	(0.00128)	(0.0276)	(0.184)	(0.00345)	(0.0242)
Female × Irish	−0.0356	−0.000928	−0.0183	−0.395	−0.00754	−0.0384
	(0.0901)	(0.00193)	(0.0415)	(0.276)	(0.00519)	(0.0364)
Number of observations	9,250	9,250	9,250	9,250	9,250	9,250
R^2	0.700	0.704	0.0362	0.589	0.612	0.0446

Source: Bowditch (1877).
Notes: Z-scores are computed relative to the modern standard using the LMS method, as incorporated in the zanthro command in Stata. The modern standard for each (sex × age) comes from the 2000 CDC Growth Charts. A full set of age dummies is included in each regression. The Bowditch tables give the physical weights in five-pound bins (e.g., 50 to 54). We used the lower bound in these calculations.
*** $p < 0.01$.
** $p < 0.05$.

we find no evidence that girls, overall, were relatively deprived in the 1870s compared with those today.

Why were girls more likely to die than boys when infectious disease was more prevalent? It is possible that young girls had greater exposure to disease because they were more likely to look after infants and toddlers, and to care for sick and aged family members when needed.[38] The possibility could be evaluated with data on disease and longevity by birth order and sex. We would expect that females who grew up with younger siblings would be more likely become infected and live shorter lives as a consequence, relative to males with the same number of younger siblings. The disadvantage would decline and disappear altogether with a less virulent pathogenic environment.

Another reason is that females relative to males become more susceptible to infectious disease sometime in childhood, despite being less susceptible as infants. The adolescent growth spurt and menstruation increase the body's protein and iron requirements. If these requirements are not met, infectious disease resistance drops.

Today anemia and micronutrient deficiencies are greater among girls than boys. Pregnancy also increases the severity of many infectious diseases such as influenza, hepatitis, herpes, and malaria.[39] Thus, in environments with poor nutrition, infectious disease immunity might decrease for women from the time of their adolescent growth spurt and continuing in their most fertile years. Improvements in nutrition would lower the overall incidence of infectious disease, and disproportionately benefit young females.[40]

5. Conclusion: relationship between early infectious disease and later death rates

We have demonstrated that young females in the US died more from infectious disease than did young males before the early twentieth century. Exactly why that was the case is not yet clear, although it does not seem to have been caused by relative deprivation. Although we have no direct evidence for this, young females must also have had greater exposure or susceptibility to infectious disease than young males—i.e., a greater morbidity rate—and carried with them, through life, the scarring effects of early illnesses. As infectious diseases were reduced, females gained more years of life as children and also as adults. We view this as a potentially important factor in the growing female life expectancy advantage.[41]

Early infectious disease can impact the mortality risk of a cohort in several ways. One is a positive selection effect: early disease culls the weak. The other is a negative scarring effect: infectious disease early in life leaves the cohort with a variety of frailties and susceptibilities. Although we do not have the incidence and prevalence of infectious disease in the cohorts we are studying, it is likely that the higher the death rate from a disease early in a cohort's life, the higher the burden of that disease is among the living. (On the 1918 flu pandemic, see Almond, 2006.)

The scarring effect probably dominated the two, at least historically (Hatton, 2001). It has been demonstrated that cohorts exhibit a "morbidity phenotype." According to Finch and Crimmins (2004), who use data for Swedish cohorts born during 1750–1940, higher levels of mortality early in life are related to higher levels of mortality later in life. In a related paper, Crimmins and Finch (2006) show that cohorts with a presumably lesser burden of infectious disease at younger ages had lower levels of adult mortality and were also taller.[42] Thus, cohorts born in times of lower infectious disease burdens are healthier, despite the fact that more survive.

Why the burden of prior infectious disease decreases longevity measures for a cohort is likely due to two effects on the survivors. The first is the long-run impact of certain infectious diseases that weaken various organs. The best known of these diseases is rheumatic fever, which damages the valves of the heart and often leads to rheumatic heart disease later in life (Elo and Preston, 1992). For some infectious diseases, the virus that sickened the person remains in the body and reappears later to trigger another disorder. Such is the case of the herpes chickenpox virus and its later-life

[38] Alsan et al. (2017) show that in low and middle income countries today, education decreases relatively for females when a younger sibling (less than five years) is ill. Consistent with this result, we find that in the 1870 census, females aged 10 to 20 years were far more likely to be engaged in occupations relating to the care of others relative to males (20 versus 1 percent). Whereas females were more likely to "keep house," and to be housekeepers, teachers, and laundresses, males were far more likely to be engaged in agricultural jobs.

[39] Kourtis et al. (2014) reviews the literature on the subject including immunologic alterations during pregnancy.

[40] On the role of nutrition in health and longevity, see the widely cited works of McKeown (1976) and Fogel (for example, Floud et al., 2011).

[41] Preston (1976), in an extensive volume on cause of death, devotes a chapter to sex differences and shows that at higher levels of mortality the female advantage disappears and females die more from almost all causes. Preston also notes the higher death rates of females than males from infectious disease from ages 1 to 30 in places of high infectious disease.

[42] We say "presumably" because they use the death rate among children at various ages rather than the death rate from infectious disease. The period they use is for cohorts born in 1900, in the era of high infectious disease.

form, shingles. Other examples include the viruses HBV and HCV, which cause liver cancer later in life; HPV, which causes cervical cancer, and bacteria *H. pylori* causing stomach cancer. There are, in addition, a host of later-life disorders that have been linked to early infections in a direct manner, such as those that weaken respiratory organs.

Associations also exist that are not as obviously connected to a damaged organ or to an infectious agent that remains dormant for some time.[43] Inflammatory indicators provide the second effect on survivors and have been related to vascular and other diseases of older age. Thus a mechanism for later disease and disability is the inflammatory response of infectious disease early in life (Crimmins and Finch, 2006; Finch and Crimmins, 2004).[44]

The relationship between infectious disease in early life and later-life health seems clear, but there are few estimates of the impact at the population level. More important with regard to our question, we know of none that distinguishes the impact on morbidity and mortality by gender, and few by the type of infectious agents.[45] Demographic historians, Bengtsson and Lindstrom (2000, 2003), show for Sweden that a higher disease burden in infancy, mainly from whooping cough and small pox as proxied by the infant mortality rate, is associated with greater mortality at ages 55 to 80, especially from airborne infectious disease. Costa (2003), using the rich Union Army data, shows that reductions in childhood infectious disease rates account for 13 percent of the increase in survival rates among 50 to 64 year old men across the twentieth century.

Our paper has uncovered (or rediscovered) an important change in the health of females in their childhood and teen years. The precise relationship between that improvement and the female longevity advantage is not yet known. But there is good reason to believe that females, more so than males, were greatly advantaged as children and as adults by the sharp reduction in infectious disease in the early twentieth century.

Acknowledgments

The authors thank the research assistants who helped assemble and analyze the data and check copy. In Cambridge, the group included: Elizabeth Engle, Celena Huo, Namrata Narain, Dev Patel, and Cesia Sanchez. In Los Angeles, the group included: Carolina Arteaga, Keyoung Lee, and Weijia Zhao. This paper was presented at the NBER Cohort Studies conference honoring the contributions of Robert W. Fogel, May 11-12, 2018.

Appendix A. Table A1: Maternal Mortality Calculation

The calculation was done for 1880 to 1990. We use life expectation at 1900, but substituting 1900 for 1880 does not change the results since maternal mortality was about the same in the two years. The calculation assumes that the fertility rate in 1880 by age is that in 1990 US and also that the maternal mortality rate (deaths from maternal causes/100,000 live births) is reduced from that in 1880 to that in 1990. The resulting decrease in deaths by age group is:

Age Group	Percentage Decrease in Deaths in 1880 or 1900 if Fertility and Maternal Mortality Decrease to 1990 Levels
15 < 20	9%
20 < 25	17%
25 < 30	14%
30 < 35	9%
35 < 40	6%
40 < 45	2%

The decrease in deaths is then applied to these age groups in computing life expectations in 1900 from data for native born whites in Haines (1998):

	Life Expectation: Additional Years Left		
Age	1900	1900 with maternal mortality decrease	Difference in years
0	51.93	52.51	0.58
1	57.46	58.10	0.64
2	58.30	58.96	0.66
3	58.16	58.83	0.67
4	57.73	58.40	0.67
5 < 10	57.15	57.84	0.69
10 < 15	53.29	53.98	0.69
15 < 20	48.94	49.65	0.71
20 < 25	44.96	45.68	0.72
25 < 30	41.35	42.06	0.71
30 < 35	37.81	38.50	0.69
35 < 40	34.24	34.89	0.65
40 < 45	30.57	31.15	0.58

[43] We should also mention the "Barker hypothesis" and the huge literature that it has spawned. According to Barker and others, nutrition in early childhood and in utero, as well as the various environmental and infectious insults experienced by the mother during the different stages of pregnancy, may contribute to increased mortality in childhood and old age (Barker et al., 2002). These shocks are hypothesized to affect the development of several organs in utero.

[44] The medical literature on the impact of inflammation on life expectancy is enormous. Blackwell et al. (2001) use self-reported data on childhood illness from the Health and Retirement Study and find strong associations between illness and poor health in middle age controlling for socio-economic status. The associations, moreover, are stronger when the childhood illness was an infectious disease. They do not explore mortality differentials.

[45] Almond et al. (2018) review the literature relating infectious diseases to later outcomes and show impacts on test scores, income, educational attainment, welfare take-up, and chronic conditions, but not mortality directly. Haas (2007) reviews existing work that links childhood illness and adult morbidity.

Appendix B Table B2: Gender Differences among Children in the Nineteenth Century US

Dependent variables	Constant	(Female mean)	Male = 1		N
Part A: 1850-1900 US Censuses					
All children under 14					
% not living in households	0.009***	[0.000]	0.001**	[0.000]	792,846
% abandoned[a]	0.003***	[0.000]	0.001***	[0.000]	792,846
% children ages 5-14 in school	0.587***	[0.001]	−0.002	[0.001]	505,501
All children under 14 living in household quarters					
% children ages 5-14 in school	0.586***	[0.001]	−0.002	[0.001]	499,939
Family size (own-family members in household)	6.44***	[0.004]	0.006	[0.005]	785,495
Number of siblings in household	3.19***	[0.004]	0.007	[0.005]	785,495
Property value of father (1850-1870)[b]	1,780***	[18.81]	−12.495	[26.4]	295,372
Property value of mother (1850-1870)[b]	95.72***	[5.548]	−6.365	[7.79]	311,466
Part B: 1850 US Census					
All children under 14					
% not living in households	0.012***	[0.001]	0.000	[0.001]	80,960
% abandoned[a]	0.002***	[0.000]	0.001**	[0.000]	80,960
% children ages 5-14 in school	0.599***	[0.003]	0.006	[0.004]	51,918
All children under 14 living in household quarters					
% children ages 5-14 in school	0.598***	[0.003]	0.006	[0.004]	51,207
Family size (own-family members in household)	6.802***	[0.013]	0.036**	[0.018]	79,977
Number of siblings	3.488***	[0.012]	0.035**	[0.017]	79,977
Property value of father	1292***	[39.27]	−34.25	[55.0]	70,541
Property value of mother	44.020***	[4.712]	−0.521	[6.60]	73,399

Sources: Ruggles et al. (2019). IPUMS, 1850, 1860, 1870, 1880 and 1900 Censuses. 1% samples.
Notes: Each line is a separate equation. Unweighted regressions are reported. Other than a male dummy, no other covariates are included in the regressions. Standard errors are in brackets.
[a] Abandoned refers to children living in correctional institutions, mental institutions or in institutions for the handicapped and poor.
[b] These variables were collected only in 1850, 1860 and 1870.
*** p<0.01, ** p<0.05, * p<0.1.

References

Albanesi, Stefania, Olivetti, Claudia, 2016. Gender roles and medical progress. J. Polit. Econ. 124 (3), 650–695.
Almond, Douglas, 2006. Is the 1918 influenza pandemic over? Long-term effects of *In-Utero* influenza exposure in the post-1940 U.S. population. J. Polit. Econ. 114 (4), 672–712.
Almond, Douglas, Currie, Janet, Duque, Valentina, 2018. Childhood circumstances and adult outcomes: act II. J. Econ. Lit. 56 (4), 1360–1446.
Alsan, Marcella, Goldin, Claudia, 2018. Watersheds in child mortality: the role of effective water and sewerage infrastructure, 1880 to 1915. J. Polit. Econ. 127 (2), 586–638.
Alsan, Marcella, Xing, Anlu, Wise, Paul, Darmstadt, Gary L., Bendavid, Eran, 2017. Childhood illness and the gender gap in adolescent education in low- and middle-income countries. Pediatrics 140 (1), 1–8.
Arnold, A.P., Chen, X., Itoh, W., 2012. What difference an X or Y makes: sex chromosomes, gene dose, and epigenetics in sexual differentiation. Handb. Exp. Pharmacol. 214, 67–88.
Austad, Steven N., 2006. Why women live longer than men: sex differences in longevity. Gend. Med. 3 (2), 79–92.
Barcellos, Silvia Helena, Carvalho, Leandro S., Lleras-Muney, Adriana, 2014. Child gender and parental investments in india: are boys and girls treated differently? Am. Econ. J. Appl. Econ. 6 (1), 157–189.
Barford, Anna, Dorling, Danny, Smith, George Davey, 2006. Life expectancy: women now on top everywhere during 2006. Even in the Poorest Countries, women can expect to outlive men. Br. Med. J. 332 (7545), 808.
Barker, D.J., Eriksson, J.G., Forsen, T., Osmond, C., 2002. Fetal origins of adult disease: strength of effects and biological basis. Int. J. Epidemiol. 31 (6), 1235–1239.
Beltran-Sanchez, Hiram, Finch, Caleb E., Crimmins, Eileen M., 2015. Twentieth century surge of excess adult male mortality. Proc. Natl. Acad. Sci. 112 (29), 8993–8998.
Bengtsson, Tommy, Lindstrom, Martin, 2000. Childhood misery and disease in later life: the effects on mortality in old age of hazards experienced in early life, Southern Sweden, 1760-1894. Popul. Stud. 54 (3), 263–277.
Bengtsson, Tommy, Lindstrom, Martin, 2003. Airborne infectious diseases during infancy and mortality in later life in southern sweden, 1766-1894. Int. J. Epidemiol. 32 (2), 286–294.
Blackwell, Debra L., Hayward, Mark D., Crimmins, Eileen M., 2001. Does childhood health affect chronic morbidity in later life? Soc. Sci. Med. 52 (8), 1269–1284.
Bowditch, Henry Pickering, 1877. The Growth of Children. Albert J. White State Printer, Boston.
Chen, X., McClusky, R., Chen, J., Beaven, S.W., Tontonoz, P., Arnold, A.P., Reue, K., 2012. The number of X chromosomes causes sex differences in adiposity in mice. PLoSGenetics 8 (5), e1002709.
Commonwealth of Massachusetts. 1887 to 1940. *Annual Report on the Vital Statistics of Massachusetts for the Year Ending December 31 [Year]*. Boston: Wright and Potter Printing Co., State Printers.
Costa, Dora., 2003. Understanding mid-life and older age mortality declines: evidence from union army veterans. J. Econom. 112 (1), 175–192.
Courtwright, David T., 1990. The neglect of female children and childhood sex ratios in Nineteenth-Century America: a review of the evidence. J. Fam. Hist. 15 (3), 313–323.
Crimmins, Eileen M., Finch, Caleb E., 2006. Infection, inflammation, height, and longevity. Proc. Natl. Acad. Sci. 103 (2), 498–503.
Cullen, Mark R., Baiocchi, Michael, Eggleston, Karen, Loftus, Pooja, Fuchs, Victor, 2016. The weaker sex? Vulnerable men and women's resilience to socio-economic disadvantage. SSM Popul. Health 2, 512–524.
Cutler, David, Miller, Grant, 2005. The role of public health improvements in health advances: the twentieth-century United States. Demography 42 (1), 1–22.
Dahl, Gordon B., Moretti, Enrico, 2008. The demand for sons. Rev. Econ. Stud. 75, 1085–1120.
Davenport, R., 1900. Annual Deaths by Cause, Age and Sex in England and Wales, 1848–1900. [data collection], 2nd Edition. UK Data Service, SN, pp. 5705 http://doi.org/10.5255/UKDA-SN-5705-1.
Doege, T.C., 1965. Tuberculosis mortality in the United States: 1900 to 1960. J. Am. Med. Assoc. 192, 1045–1048.
Drevenstedt, Greg L., Eileen Crimmins, M., Vasunilashorn, Sarinnapha, Finch, Caleb E., 2008. The rise and fall of excess male infant mortality. Proc. Natl. Acad. Sci. 105 (13), 5016–5021.
Du, S., Itoh, N., Askarinam, S., Hill, H., Arnold, A., Voskuhl, R., 2014. XY sex chromosome complement, compared with XX, in the CNS confers greater neurodegeneration during experimental autoimmune encephalomyelitis. Proc. Natl. Acad. Sci. 111 (7), 2806–2811.
Elo, Irma T., Preston, Samuel H., 1992. Effects of early-life conditions on adult mortality: a review. Popul. Index 58 (2), 186–212.
Enterline, Philip E., 1961. Causes of death responsible for recent increases in sex mortality differentials in the United States. Millbank Mem. Fund Q. 39 (2), 312–328.
Finch, Caleb E., Crimmins, Eileen M., 2004. Inflammatory exposure and historical changes in human life-spans. Science 305, 1736–1739.
Floud, Roderick, Fogel, Robert W., Harris, Bernard, Hong, Sok Chul, 2011. The Changing Body: Health, Nutrition, and Human Development in the Western World Since 1700. Cambridge University Press, Cambridge.

Goldin, Claudia, Katz, Lawrence F., 2008. The Race Between Education and Technology. Belknap Press, Cambridge, MA.
Haas, Steven A., 2007. The long-term effects of poor childhood health: an assessment and application of retrospective reports. Demography 44 (1), 113–135.
Hacker, J.David., 2010. Decennial life tables for the white population of the United States, 1790-1900. Hist. Methods A J. Quant. Interdiscip. Hist. 43 (2), 45–79.
Haines, Michael., 1998. Estimated life tables for the United States, 1850–1910. Hist. Methods J. Quant. Interdiscip. Hist. 31 (4), 149–169.
Hammel, E.A., Johansson, Sheila R., Ginsberg, Caren A., 1983. The value of children during industrialization: sex ratios in childhood in Nineteenth-Century America. J. Fam. Hist. 8 (4), 346–366.
Hatton, Timothy J., 2001. Infant mortality and the health of survivors: Britain, 1910-50. Econ. Hist. Rev. 64 (3), 951–972.
Henry, Louis., 1989. Men's and women's mortality in the past. Popul.: Engl. Selection 44 (1), 177–201.
Historical Statistics, 2006. Historical Statistics of the United States, millenial edition.
Human Mortality Database, 2019. Human Mortality Database. University of California, Max Planck Institute for Demographic Research (Germany), Berkeley (USA), Available at www.mortality.org or www.humanmortality.de (data downloaded 8/24/2017).
Jayachandran, Seema, Lleras-Muney, Adriana, Smith, Kimberly V., 2010. Modern medicine and the twentieth century decline in mortality: new evidence on the impact of sulfa drugs. Am. Econ. J. Appl. Econ. 2 (2), 118–146.
Jayachandran, Seema, Pande, Rohini, 2017. Why are Indian children so short? The role of birth order and son preference. Am. Econ. Rev. 107 (9), 2600–2629.
Jensen, Robert., 2002. Fertility Preferences and Female Disadvantage: Equal Treatment, Unequal Outcomes? Mimeo.
Johansson, Sheila Ryan., 1984. Deferred infanticide: excess female mortality in childhood. In: Hausfatter, Glenn, Hrdy, Sarah Blaffer (Eds.), Infanticide: Comparative and Evolutionary Perspectives. Aldine Press, New York.
Johnston, William, 1995. The Modern Epidemic: A History of Tuberculosis in Japan. Harvard University Press.
Klasen, Stephan, 1998. Marriage, bargaining, and intrahousehold resource allocation: excess female mortality among adults during early german development 1740–1860. J. Econ. Hist. 58 (2), 432–467.
Kourtis, Athena P., Read, Jennifer S., Jamieson, Denise J., 2014. Pregnancy and infection. N. Engl. J. Med. 370, 2211–2218.
Logan, Trevon., 2007. On Family Allocation Strategy in the Late Nineteenth Century. Ohio State Working Paper.
Loudon, Irvine., 1992. Death in Childbirth: An International Study of Maternal Care and Maternal Mortality 1800-1950. Oxford University Press, New York.
Lynch, Katherine A., Greenhouse, Joel B., 1994. Risk factors for infant mortality in nineteenth-century Sweden. Popul. Stud. 48 (1), 117–133.
Madigan, Francis C., 1957. Are sex mortality differentials biologically caused? Millbank Mem. Fund Q. 35 (2), 202–223.
McKeown, Thomas, 1976. The Modern Rise of Population. Academic Press, New York.
Noymer, Andrew, Garenne, Michel, 2000. The 1918 influenza epidemic's effects on sex differentials in mortality in the United States. Popul. Dev. Rev. 26 (3), 565–581.
Omran, Abdel R., 2005. The epidemiologic transition: a theory of the epidemiology of population change. Milbank Q. 83 (4), 731–757.
Pinnelli, Antonella, Mancini, Paola, 1997. Gender mortality differences from birth to puberty in Italy, 1887-1940. In: Corsini, Carlo A., Viazzo, Pier Paolo (Eds.), The Decline of Infant and Child Mortality: The European Experience, 1750-1990. Klewer Law International, The Hague.
Preston, Samuel H., 1976. Mortality Patterns in National Populations: With Special Reference to Recorded Causes of Death. Academic Press, New York.
Preston, Samuel H., Wang, Haidong, 2006. Sex mortality differences in the United States: the role of cohort smoking patterns. Demography 43 (4), 631–646.
Retherford, Robert D., 1972. Tobacco smoking and the sex mortality differential. Demography 9 (2), 203–215.
Ruggles, Steven, Flood, Sarah, Goeken, Ronald, Grover, Josiah, Meyer, Erin, Pacas, Jose, Sobek, Matthew, 2019. IPUMS USA: Version 9.0 [dataset]. IPUMS, Minneapolis, MN, http://dx.doi.org/10.18128/D010.V9.0.
Smith, Nicole L., 2008. The Problem of Excess Female Mortality: Tuberculosis in Western Massachusetts, 1850-1910. Unpublished Masters Thesis. University of Massachusetts, Amherst.
Stolnitz, George J., 1956. A century of international mortality trends: II. Popul. Stud. 10 (1), 17–42.
U.S. Department of the Interior. Census Office, 1890. Population of the United States. Part I. U.S. GPO, Washington, D.C.
U.S. Department of the Interior. Census Office, 1900. Population of the States and Territories, Vol. I. U.S. GPO, Washington, D.C.
U.S. Department of Commerce. Bureau of the Census, 1910. Population. General Report and Analysis, Vol. I. U.S. GPO, Washington, D.C.
U.S. Department of Commerce. Bureau of the Census, 1920. Population. Number and Distribution of Inhabitants. U.S. GPO, Washington, D.C.
U.S. Department of Commerce. Bureau of the Census, 1930. Population. Number and Distribution of Inhabitants, Vol. I. U.S. GPO, Washington, D.C.
U.S. Department of Commerce. Bureau of the Census, 1940. Population. Number of Inhabitants, Vol. I. U.S. GPO, Washington, D.C.
U.S. Social Security Administration, 2005. Life tables for the social security area, 1900-2100. In: Bell, Felicitie C., Miller, Michael L. (Eds.), Actuarial Study No. 120. , August https://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl_6_1900.html.
U.S. Social Security Administration, 2017. Period Life Tables. On-line Data Updating Actuarial Study No. 120. https://www.ssa.gov/oact/HistEst/PerLifeTables/2017/PerLifeTables2017.html.
Zeng, Yi, Ping, Tu, Baochang, Gu, Yi, Xu, Bohua, Li, Yongping, Li, 1993. Causes and implications of the recent increase in the reported sex ratio at birth in China. Popul. Dev. Rev. 19 (2), 283–302.