For Presentation at the Conference Honoring Robert W. Fogel
NBER, Cambridge MA May 11 & 12, 2018
April 6, 2018, version
Comments Welcome

XX > XY?: The Changing Female Advantage in Life Expectancy

Claudia Goldin
Harvard University
NBER

Adriana Lleras-Muney
University of California at Los Angeles
NBER

Acknowledgments: The authors thank the many research assistants who helped assemble and analyze the data and check copy. In Cambridge, the group included: Elizabeth Engle, Celena Huo, Namrata Narain, and Cesia Sanchez. In Los Angeles, the group included: Carolina Arteaga, Keyoung Lee, and Weijia Zhao.

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. Many of the reasons why women live longer than men can be found in an understanding of the role of having two X chromosomes, rather than an X and a Y, and the resulting impact on hormones.[3] Whereas some biological factors are at play, the fact that the advantage has increased greatly in the latter half of the twentieth century, suggests that environmental factors have disproportionately benefited women. In this paper we examine *when* the female advantage started to grow to understand *what* caused the advantage to widened.

We do not have a complete answer concerning why there was a large increase in the female advantage and many factors certainly contributed to it. We argue that the reduction in infectious diseases as a major cause of death played an important role. We use new data from Massachusetts containing information on causes of death from 1887 to show that infectious diseases disproportionately affected females between the ages of 5 and 25. Increased longevity of women, therefore, occurred as the burden of infectious disease fell

[1] See, for example, Barford et al. (2006), although the data are not entirely reliable for certain developing countries (and are dismal for many, 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 above 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 above 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 for which males are four to ten times more likely to be affected. Females are more likely to get autoimmune diseases, but their extra X chromosome protects them from more rapid degeneration once affected.

for all. Our explanation does not tell us why women live longer than men, but it does help in understanding the timing of the increase.

Mortality rates have always been higher for male than female infants (Drevenstedt et al. 2008). But, even conditional on surviving to year one, females had about the same expectancy, or somewhat greater, than did males by the latter part of the nineteenth century.[4] The advantage, though, pales in comparison to what would soon emerge. In today's developed countries gender gaps in longevity rose to a maximum of eight years in the 1970s and the difference then narrowed (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 estimates are 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 embraced by women as well and that change 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 explain the rise, and fall, in longevity gender gaps (Cullen et al. 2016; Preston and Wang 2006).

---

[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 gap at birth increased by 3.43 years 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 1, is around 15 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 Wong 2006).

Our contribution is to explore the reasons for the initial appearance and widening of the female longevity advantage. In the US, as well as in England and 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.

Young females initially 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 due to differential treatment by sex. 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 appeared as an even greater 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. Although our data are mainly in period, not cohort, form, our work informs a cohort analysis presented at the end of the paper. 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 expectation 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 the later-life mortality.

I. Life Expectation for Males and Females: The Long Run

A. The US: 1795 to 2015

Life expectation at birth (e_0) increased greatly from the late nineteenth century—when it was around 50 for females—to the mid-twentieth century—when it was around 70 years 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 expectation series for the US in Figure 1.1, the female advantage is clearly evident around the 1880s to the 1890s, after which the difference between the sexes expanded. The widening intensified in the first third to first half of the twentieth century, and particularly in the post WWII period, when the growth in male life expectancy decelerated considerably. These trends reversed course in the 1970s when male life expectancy started to grow faster again, and later in the 1990s when female life expectancy growth decelerated. The longevity gender gap has narrowed since and is about five years today.

The first of the Figure 1.1 graphs (part A) gives life expectation at birth. Until the 1890s life expectation at birth was about equal by sex. Part B, giving expectation at age 15, shows a male advantage until the 1890s. The final graph (part C), giving expectation conditional on reaching at 45, shows a female advantage present throughout the period, albeit initially quite small. It should be noted that in the cases of parts B and C longevity is expressed as expected age at death, not as years left to live. Based on this information it would appear that the female advantage at birth began in earnest in the US in the 1890s.

The shaded portion in each of the Figure 1.1 graphs gives the period of splicing of two series, that from J. David Hacker (2010) for the period to 1900 and that from the US Social Security Administration (SSA) for the years after. The Hacker series is always a bit higher than that from SSA probably because the Hacker series is for whites. Another important feature of the graphs is the sudden and short-lived disadvantage for males

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

around 1865, due to excess male mortality from the US Civil War.

Trends in the female survival advantage can be seen better in Figure 1.2, which gives the difference (male minus female) in life expectation 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 decade of the 1910s due to World War I. The opposite impact around 1925 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, war casualties overwhelmed it early on.

The conclusions from Figures 1.1 and 1.2 are that the female advantage at age 15 appeared around the turn of the twentieth century, whereas that at birth and conditional on surviving to age 45, they were always present to a small degree until the late nineteenth century. That said, the advantage widened considerably in all three cases after 1930 and then narrowed somewhat after 1970.

Comparing the data on life expectation at age 15 (e_{15}) with that at age 45 (e_{45}) shows another important trend. 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 even throughout the nineteenth century. Thus the mortality consequences of childbirth, either due to maternal mortality or to the later mortality consequences of maternal morbidity, are a likely reason for the differences between e_{15} and e_{45}. But we will suggest that there may be other reasons why women began to live a lot longer than men starting in the late nineteenth century.

B. Europe: Early-Nineteenth Century to 2000s

We provide similar life expectancy data and differences by sex at various ages for England and 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, females in these nations had an advantage earlier than in the US. Recall that in the US, the female

advantage appeared in earnest around the 1890s.

For the European data we use the Human Mortality Database (HMD) and present the data in Figure 1.3 as the difference between male and female life expectation at birth and conditional on surviving to ages 15 and 45 (e_0, e_{15} and e_{45}) as we did in Figure 1.2 for the US.[8] Females in England (part A) and France (part B) had a significant edge on males from birth from the mid- to late-nineteenth century and the advantage then greatly expanded as it did in the US.[9] Sweden (part C) is different. A female mortality advantage is apparent at all ages and as far back as the data will allow us to go.[10] But these differences *shrink* up to 1930. In all the countries considered, including the US, the female advantage greatly expanded after the WWII and then started to shrink after the 1970s.

The reason for "Swedish exceptionalism" is not entirely clear but may stem from the low prevalence of infectious disease in nineteenth century Sweden. Swedish infant mortality rates were low through the nineteenth century. They started their decline 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 advatange emerged earlier in Sweden because the infectious disease environment was lower.[11] But no data on mortality by cause of death exist to confirm the hypothesis. We return later to the important role of infectious disease.

---

[8] 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.

[9] 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).

[10] We only plot years with population censuses. Longer time series for Sweden have been constructed using alternative sources for population counts.

[11] 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.

Our discussion of the European mortality data has ignored the role of war largely because our figures include only years when census data was collected (and thus when we have accurate population counts) and censuses are not taken during conflict. 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 the number of 20<30 year old men killed in the 1914-1919 period was 10 times larger than that for females (15 times larger in France).[12] 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.

II. Sex Differences in Mortality among Children and Young Adults

A. Relative Deaths by Sex: Massachusetts

Historical evidence on deaths by age and cause of death for the state of Massachusetts (MA) allow us to better understand aggregate US trends. MA has the longest and highest quality vital statistics data among all US states.[13] Death counts broken down by sex, age-groups, and cause of death are available annually starting in 1885. Identical data for the entire US population are available only starting with 1933, well after the female advantage appears.

Figure 2.1, 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 20<30 years.[14] Because of the relatively small number of annual deaths in some of these age brackets, the data are expressed as a three-year centered moving average. We exclude 1918 from the three-year average due to the large number of deaths from the pandemic but include it separately 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 line of interest is the dashed line giving the

[12] Results not shown but available upon request.
[13] Massachusetts, in 1842, was the first state to pass a state registration law.
[14] We begin the series with 1887 because the probable cause of death was aggregated in some meaningful manner beginning in that year.

ratio (Males/Females) of the two series. The ratio is mapped on to the right axis and an arrow points to equality.

The ratio of male to female deaths becomes greater than one around 1895 to 1900 in all four cases. More important, there is a distinct decrease in the male to female ratio for the 10<15 year olds and a smaller one for the 5<10 year olds around 1918. That is, around 1918 females 10<15 were dying at a higher rate than were males. It is possible that these differences extend, as well, 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 below one for the 20<30 year olds starting in 1918 and lasting until 1935.[15]

The main conclusion is that female children, especially those 10<15 years old, died at higher rates in 1918 from the flu than did males of the same age group and relative to previous rates by sex. Note that relative deaths after the flu pandemic resumed their previous trend of an increasing female advantage.

The fact that more females 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.[16] The answer we will provide is that they did. 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.[17]

---

[15] 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.
[16] Young women (less than 25 years) succumbed to pulmonary tuberculosis at higher rates than did young men (Hacker 2010; Henry 1989; Madigan 1957). The opposite is true among adults older than 25 years—males appear more susceptible to TB (Noymer and Garenne 2000). 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., Drevenstadt et al. 2008). But after around age five, females appear to have fallen prey to the various infectious diseases more than did males.
[17] 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

Of equal interest is that the 20<30 year old group in MA shows a distinct break around 1920. Previously, women were dying at lower rates than men in that age group. But after 1920 both men and women in that age group died at significantly lower levels. 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 from childbirth, however, had not yet decreased and therefore the remaining deaths in the 20<30 year old group tilted toward females.

B. Relative Deaths by Sex: England and France

The death data by sex in both England and France yield similar findings. Surprisingly in this case, the female advantage among youths 10<15 appears later in England and France than in MA.

Recall that life expectation at various ages demonstrated an earlier female advantage than in the US. But as can be seen in Figure 2.2, deaths among male youths 10<15 in England systematically started to exceed deaths of female youths 10<15 only sometime around 1925, and they never exceeded female deaths in France. Figure 2.2 has been drawn in a similar manner to that of Figure 2.1 for MA.[18] In a parallel manner to the MA case, girls suffered greater losses than did 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 since they would have been too young to have served. Because 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 the effects of the flu.

C. Deaths Rates by Sex: MA and Europe

---

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.

[18] 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.

The main finding from Figure 2.1 was that the 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 are not expressed as a fraction of the population. The population of males and females by age deviated from equality because in MA young men left the state in the 1920s more than did women, as the manufacturing economy of the Commonwealth was tanking.[19] We now present data for census years that divides by population and expresses these numbers as rates for census years.

We 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.[20] In between census years accurate population estimates can only be computed with information on migration, which does not exist at high enough frequency by age and sex.

Figure 2.3 provides the 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

[19] The ratio of males to females in the 20<30 year group declined in the 1910s and 1920s and then recovered in the 1930s. Source: US Census of Population, Massachusetts.

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

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

prime reproductive ages would become substantially less than that for men.[21]

Figure 2.4 (drawn to the same scale as Figure 2.3) shows similar mortality rate data for England from roughly 1880 to 1930. The female advantage for these four ages groups 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 mortality advantage of females below age 20 in MA was about 1.2 or higher but it was never above 1.2 in England.

D. 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 the 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 1900, to 14 in 1901, 15 in 1921 and 18 in 1931.

A far greater fraction of deaths to females than to males in the 10<15 year group were due to infectious disease throughout the period, as can be seen in Figure 2.5, part A. But more males than females died from external and violent causes. Therefore, we must also show that deaths from infectious disease were greater for females than males even if we exclude the violent and external death category. Except for the years after 1925 and for those with significant infectious disease epidemics, the fraction of non-violent deaths from infectious disease was higher for females than for males, as can be seen in Figure 2.5, 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

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

decreased more for females than males.[22]

The data in Figure 2.5, like those in Figure 2.1, are not expressed as rates. To demonstrate the decreasing importance of infectious disease as a cause of death, we must divide by the appropriate population. In Figure 2.6 we express deaths as a fraction of the age and sex relevant population groups for census years and use only deaths from non-violent factors for the 10<15 year group.[23] A clear decrease in the death rate can be observed for both males and females in the 1890 to 1940 period, although the decrease is larger for females. Because infectious disease was a greater factor in female than male deaths, as it declined as a cause of death, females were disproportionately advantaged.

E. 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 US maternal mortality series begins in 1915, but the MA series can start in the 1870s with vital statistics reports that provide sufficient detail on cause of death. We produce a maternal mortality series for MA defined as the number of deaths from maternal causes per 100,000 live births, given in Figure 2.7, using death and birth records.[24] 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 in the period of the flu pandemic, after which it fell beginning in the early 1930s. The MA series is generally lower than that for the US,

[22] 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.
[23] 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 infectious elements, although some may have been described as due to diseases of certain organs.
[24] See Figure 2.8 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.

but still considerable higher than in the England or Sweden in 1930.[25]

Because the general death rate decreased in the 1890s as infectious disease began to wane, but that due to maternal mortality did not decrease until much later and rose for some of the period, the fraction of female deaths due to maternal causes increased, as is shown in Figure 2.8. The fraction of all deaths to women 20<30 years old from maternal causes was less than 6 percent in 1888 but was almost 18 percent—three times—in 1930 (see dotted line graphed to the right axis). After the advent of sulfa drugs, deaths due to maternal causes also fell and the fraction due to those causes began its descent, as can be seen at the end of the series.

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 think 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, the fraction of the rising female advantage due to the decrease in maternal mortality was only about a seventh. In part this is because even for this fertile age group TB, not maternal mortality, was the leading cause of death for females in 1930 (Enterline 1961).[26]

We can show this as follows. If maternal mortality rates in 1900 were decreased to that existing in 1990—almost zero—life expectancy for women in the US would have increased by about 0.6 years at birth and 0.7 years at age 20 (see Appendix Table 1). That is not zero, of course. But the gain from eliminating maternal mortality pales in comparison with the actual increase in life expectations from 1900 to 1990 (see Figure 1.1): 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).[27]

More to the point of explaining the growing female advantage, the gap between males and females at birth was about 3 years in 1900 but was 7 years in 1990. Therefore,

---

[25] 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.
[26] Moreover TB killed more females than males in the US for those 5 to 25 years old (Doege 1965).
[27] Preston (1976) does a similar calculation (see table 3.3 and related discussion).

the decrease in maternal mortality of around 0.6 years would explain 15 percent of the increased female advantage at birth. That is certainly an overestimate since the medical advances that greatly reduced maternal mortality—antibiotics, blood transfusions, antiseptic operating, and advanced surgical techniques—would also have increased male life expectancy as well as that for females in general.

III. 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 expectation 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 due to poorer nutrition they 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 analyses of ours and 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 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 from childhood to mid-life. Stolnitz correctly emphasized several misconceptions about gender differences in mortality. The first was that females always outlive males. And the second was that the only period in which females do not

have an advantage is during the child bearing ages.[28]

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 source 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 weight of the evidence is 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—that parents want to have boys rather than girls—is weak. Hammel, Johansson and Ginsberg (1983) claim that differences in sex ratios by place show that girls were more valued where their wages were higher. But Courtwright (1990) argues that the slight imbalance of sex ratios (west having more male children than east) 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 help resolve the issue of son preference. To assess whether parents have son preference, we follow the intuition from Barcellos et al. (2014) and investigate the sex ratio by age among the youngest in the

[28] 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."

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.

Figure 3.1 shows sex ratios (defined as the number of males divided by the number of females) by age for different subsets of the population of children ages 0 to 14 years. A sex ratio at birth around 105 males per 100 females is considered by demographers to be normal. Although the data are noisy, the sex ratio does not systematically increase with age among the youngest.[29] The highest sex ratio we observe is 107. By comparison, in India the sex ratio among the youngest reaches 138 by age 4 (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 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 children's ages. The inference would be incorrect if there are children who died or moved away, but it provides a useful approximation. We do know the sex ratio in families by number of existing children. Table 3.2 shows sex ratios by observed birth order among children 0 to 20 years, and also among those ages 0 and 1, for the combined population censuses 1850 to 1900 and for 1850 alone. In the nineteenth century US, sex ratios did not greatly or systematically increase with birth order.[30]

We also investigate other indicators of son preference and find no evidence for it in

[29] 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.

[30] 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.

the US data.[31] In countries with son preference, girls are abandoned at a greater rate than are 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 2).

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.

Even in the absence of son preference, parents could treat boys and girls differently, and potentially discriminate against girls, possibly causing them to be less healthy and die at greater rates. The historical evidence on differences in family expenditures on sons and daughters is weaker and we could find none giving food allocations.[32] 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 terms of schooling, an important investment parents and societies make, there were no statistically significant differences in the rate at which boys and girls (ages 5 to 14) attended school in 1850 and 1880. And girls attended and graduated from high school in the early part of the twentieth century at a far greater rate than did boys.[33]

To evaluate the proposition that female children experienced relative nutritional deprivation in poor households, we use anthropometric evidence from Bowditch (1877), a

[31] 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.
[32] 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.
[33] On the schooling of 5 to 14 years olds in 1850 and 1880, see Goldin and Katz (2008, figure 4.2, 153). On attendance and graduation from high school, see Goldin and Katz (2008, figure 6.5, 231). Appendix Table 2 also presents results for the 1850 Censuses alone and for 1850-1900, showing no economically or statistically significant differences by gender in school attendance.

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 have been reported separately. There were, however, 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 additional difficulties 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 below, 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, where the means and standard deviations are for the native-born parentage group.[34] Only children five to ten years old can be analyzed, since the Irish sample is small at older ages probably because many of the children did not attend school. Age dummies, θ, are included. We provide the results in Table 3.3.

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

We should note that all the children in the Bowditch sample were short and thin by modern standards (the means of the zcores are negative, indicating the children were

[34] We use the definition of a Z-score as: ([X – mean]/standard deviation), which is the standard in statistics. A standard measure in anthropometrics is to use the median, as in Barcellos, Carvalho, and Lleras-Muney 2014 who use ([X – median]/standard deviation). We do not find any substantive changes when we use the median measure.

below the modern standard).[35] 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 were the native-born children. But the girls of Irish parentage were not much shorter or lighter relative to the native-born children. Although the coefficients on the interaction terms are negative, the standard errors are large and the magnitudes are small. Moreover girls, in the entire sample, were not significantly shorter or lighter in weight than males, relative to modern standards (the coefficient on female is not statistically significant in the regressions that use the z-scores as outcomes).[36]

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

IV. 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. Young females must also have had greater exposure to infectious disease than young males, that is 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

---

[35] Relative to modern growth charts at the median, the native-born parentage boys were about two inches shorter from five to seven years and two and a half inches shorter from eight to ten years old. The native-born parentage girls were as tall as the boys at the median but were an inch shorter at nine and ten years old. Relative to modern weight charts at the median, the native parentage male children were about the same weight to age eight. But were two pounds lighter from eight to ten years. Whereas modern male and female children have about the same weight from five to ten years old at the median, the Bowditch females were the same weight as their male counterparts until age eight when they were lighter by four pounds. Modern standards are from the CDC: https://www.cdc.gov/growthcharts/charts.htm#Set2

[36] The z-score normalizes the height and weight using modern populations as the reference. The insignificant female coefficient indicates that although, on average, females, were shorter and lighter compared with males (see cols. 1, 2 and 4 and 5), the gender gap is the same as today.

female life expectancy advantage.[37]

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.

The scarring effect probably dominated the two, at least historically (Hatton 2011). It has been demonstrated that cohorts exhibit a "morbidity phenotype." According to Finch and Crimmins (2004), who use data for Swedish cohorts born from 1750 to 1940, higher levels of early mortality are related to higher levels of later mortality. 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.[38] Thus cohorts born in times of lower infectious disease burdens are healthier, despite the fact that more survive. We extend the Crimmins and Finch (2006) analysis by using actual infectious disease mortality at younger ages, exploring differences in the relationship by gender, and including birth cohorts that experienced decreased infectious disease.

To explore the relationship between early-life infectious disease mortality and later-life death rates by cohort, we connect data on birth cohorts in our MA sample to death rates for the same birth cohorts in US vital statistics for 1960, 1980, and 1990.[39] Figure 4.1 graphs the relationship between the four-year mortality rate from infectious disease by cohort, age, and sex for the MA sample against the mortality rate at older ages for

[37] 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.
[38] 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 to 1900, thus in the era of high infectious disease.
[39] Death certificates contain state of birth only for 1960-1963 (in a restricted-access sample) and again starting in 1979.

individuals born in MA.[40] Each point in Figure 4.1 is for a birth cohort and age. We truncate at older ages because of small samples in vital statistics and include ages to 71 years.

For the 1960 dots, for example, individuals born in MA in 1900 are 60 years old in 1960. Their death rate in that year is given by the vertical axis and their infectious disease mortality rate, when their cohort was 10<15 years, is given by the horizontal axis. The relationship is similar when we use other ages or exclude deaths due to violent causes.

The underlying notion is that a higher infectious disease rate at younger ages also means a higher morbidity rate and thus a higher rate of scarring. The more scarring, the more the late-life burden of disease. For both males and females there is a clear positive relationship. The higher the infectious disease death rate at younger ages, the higher is the later mortality rate. The range of death rates at older ages for males is considerably greater than for females because the period covered is after the expansion of the female advantage. Although it would be better to have the later-life mortality rate data for 1940 and 1950, the data that exists begins in 1960. This exercise is a good descriptive way of visualizing the mechanism we have proposed, but it is not causal.[41]

Why the burden of prior infectious disease increases lifetime mortality for a cohort is likely due to two effects. The first is the direct impact of certain infectious diseases that weaken various organs. The best known of these is rheumatic fever, which damages the valves of the heart and often leads to rheumatic heart disease later in life (Elo and Preston 1992). In other cases, the virus remains in the body and reappears to cause another disorder, as in the case of the herpes chickenpox virus and, its later-life form, shingles. 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.

But there are also associations that are not as obviously connected to a damaged

[40] These rates are computed for each birth cohort as (Number of Deaths 60-63)/(1960 Census Population). They can be thought of as the probability of dying within four years of the census.
[41] For that we would need variation in the early disease environment by cohort, possibly by using data for several states. The problem is that few states had early vital statistics. The analysis in Crimmins and Finch (2006) is also descriptive, not causal, and it ignores important trends.

organ or to an infectious agent.[42] Demographic historians, Bengtsson and Lindstrom (2000), show for Sweden that a higher disease burden in infancy, proxied by the infant mortality rate, is associated with greater mortality from airborne infectious disease at ages 55 to 80. Inflammatory indicators have been related to vascular and other diseases of older age. Thus a probable mechanism for later disease is the inflammatory response of infectious disease early in life (Crimmins and Finch 2006; Finch and Crimmins 2004).

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 widening 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 reduction in infectious disease.

[42] 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 varous environmental and infectious insults experienced by the mother during the various stages of pregnancy, may contribute to increased mortality in childhood and old age (Barker et al. 2002).

References

Alsan, Marcella, and Claudia Goldin. 2018. "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880 to 1915," *Journal of Political Economy*, forthcoming.

Arnold, A. P., X. Chen, and W. Itoh. 2012. "What Difference an X or Y Makes: Sex Chromosomes, Gene Dose, and Epigenetics in Sexual Differentiation." *Handbook of Experimental Pharmacology* (214): 67-88.

Austad, Steven N. 2006. "Why Women Live Longer than Men: Sex Differences in Longevity," *Gender Medicine* 3(2): 79-92.

Barcellos, Silvia H., Leandro Carvalho, and Adriana Lleras-Muney. 2014. "Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics* 6(1): 157-89.

Barford, Anna, Danny Dorling, and George Davey Smith. 2006. "Life Expectancy: Women Now on Top Everywhere during 2006. Even in the Poorest Countries, Women Can Expect to Outlive Men," *BMJ: British Medical Journal* 332.7545: 808.

Barker, D.J., J.G. Eriksson, T. Forsen, and C. Osmond. 2002. "Fetal Origins of Adult Disease: Strength of Effects and Biological Basis," *International Journal of Epidemiology* 31(6): 1235-39.

Beltran-Sanchez, Hiram, Caleb E. Finch, and Eileen M. Crimmins. 2015. "Twentieth Century Surge of Excess Adult Male Mortality," *Proceedings of the National Academy of Sciences* 112(29): 8993-98.

Bengtsson, Tommy, and Martin Lindstrom. 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," *Population Studies* 54(3): 263-77.

Bowditch, Henry Pickering. 1877. *The Growth of Children*. Boston: Albert J. White State Printer.

Chen, X., R. McClusky, J. Chen, S. W. Beaven, P. Tontonoz, A. P. Arnold, and K. Reue. 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.

Courtwright, David T. 1990. "The Neglect of Female Children and Childhood Sex Ratios in Nineteenth-Century America: A Review of the Evidence," *Journal of Family History* 15(3): 313-23.

Crimmins, Eileen M., and Caleb E. Finch. 2006. "Infection, Inflammation, Height, and Longevity," *Proceedings of the National Academy of Sciences* 103(2): 498-503.

Cullen, Mark R., Michael Baiocchi, Karen Eggleston, Pooja Loftus, and Victor Fuchs. 2016. "The Weaker Sex? Vulnerable Men and Women's Resilience to Socio-Economic Disadvantage," *SSM – Population Health* 2: 512-24.

Cutler, David, and Grant Miller. 2005. "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States," *Demography* 42(1): 1-22.

Dahl, Gordon B., and Enrico Moretti. 2008. "The Demand for Sons," *Review of Economic Studies* 75: 1085–120.

Davenport, R. 2011. *Annual Deaths by Cause, Age and Sex in England and Wales, 1848-1900*. [data collection]. *2nd Edition.* UK Data Service. SN: 5705, http://doi.org/10.5255/UKDA-SN-5705-1

Drevenstedt, Greg L., Eileen Crimmins M., Sarinnapha Vasunilashorn, and Caleb E. Finch. 2008. "The Rise and Fall of Excess Male Infant Mortality, *Proceedings of the National Academy of Sciences* 105(13): 5016-21.

Du, S., N. Itoh, S. Askarinam, H. Hill, A. Arnold, and R. Voskuhl. 2014. "XY Sex Chromosome Complement, Compared with XX, in the CNS Confers Greater Neurodegeneration during Experimental Autoimmune Encephalomyelitis," *Proceedings of the National Academy of Sciences* 111(7): 2806-11.

Elo, Irma T., and Samuel H. Preston. 1992. "Effects of Early-Life Conditions on Adult Mortality: A Review," *Population Index* 58(2): 186-212.

Finch, Caleb E., and Eileen M. Crimmins. 2004. "Inflammatory Exposure and Historical Changes in Human Life-Spans," *Science* 305: 1736-39.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology*. Cambridge, MA: Belknap Press.

Hacker, J. David. 2010. "Decennial Life Tables for the White Population of the United States, 1790-1900," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 43(2): 45-79.

Haines, Michael. 1998. "Estimated Life Tables for the United States, 1850–1910," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 31(4): 149-69.

Hammel, E.A., Sheila R. Johansson, and Caren A. Ginsberg. 1983. "The Value of Children during Industrialization: Sex Ratios in Childhood in Nineteenth-Century America," *Journal of Family History* 8(4): 346-66.

Harris, Bernard. 1998. "Gender, Height and Mortality in Nineteenth- and Twentieth Century Britain: Some Preliminary Reflections." In J. Komlos and J. Baten (eds.), *The Biological Standards of Living in Comparative Perspective*. Stuttgart: Steiner: 413-48.

Hatton, Timothy J. 2001. "Infant mortality and the health of survivors: Britain, 1910-50," *The Economic History Review* 64(3): 951-72.

Henry, Louis. 1989. "Men's and Women's Mortality in the Past," *Population: An English Selection* 44(1): 177-201.

Historical Statistics of the United States, Millenial Edition. 2006.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded 8/24/2017).

Jayachandran, Seema, Adriana Lleras-Muney and Kimberly V. Smith. 2010. "Modern Medicine and the Twentieth Century Decline in Mortality: New Evidence on the Impact of Sulfa Drugs," *American Economic Journal: Applied* 2(2):118-46.

Jayachandran, Seema, and Rohini Pande. 2017. "Why Are Indian Children So Short? The Role of Birth Order and Son Preference," *American Economic Review* 107(9): 2600-29.

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 Glenn Hausfatter and Sarah Blaffer Hrdy, eds., *Infanticide: Comparative and Evolutionary Perspectives*. New York: Aldine Press.

Klasen, Steven. 1998. "Marriage, Bargaining, and Intrahousehold Resource Allocation: Excess Female Mortality among Adults during Early German Development 1740–1860," *Journal of Economic History* 58(2): 432-67.

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*. New York: Oxford University Press.

Lynch, Katherine A., and Joel B. Greenhouse. 1994. "Risk Factors for Infant Mortality in Nineteenth-Century Sweden," *Population Studies* 48(1): 117-33.

Madigan, Francis C. 1957. "Are Sex Mortality Differentials Biologically Caused?" *The Millbank Memorial Fund Quarterly* 35(2): 202-23.

Noymer, Andrew, and Michel Garenne. 2000. "The 1918 influenza Epidemic's Effects on Sex Differentials in Mortality in the United States," *Population and Development Review* 26(3): 565-81.

Omran, Abdel R. 2005. "The Epidemiologic Transition: A Theory of the Epidemiology of Population Change," *The Milbank Quarterly* 83(4): 731-57.

Pinnelli, Antonella, and Paola Mancini. 1997. "Gender Mortality Differences from Birth to Puberty in Italy, 1887-1940." In *The Decline of Infant and Child Mortality: The European Experience, 1750-1990*, Carlo A. Corsini, Pier Paolo Viazzo, eds. The Hague: Klewer Law International.

Preston, Samuel. H. 1976. *Mortality Patterns in National Populations: With Special Reference to Recorded Causes of Death*. New York: Academic Press.

Preston, Samuel H., and Haidong Wang. 2006. "Sex Mortality Differences in the United States: The Role of Cohort Smoking Patterns," *Demography* 43(4): 631-46.

Retherford, Robert D. 1972. "Tobacco Smoking and the Sex Mortality Differential," *Demography* 9(2): 203-15.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 7.0* [dataset]. Minneapolis, MN: University of Minnesota, 2017.

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. 1956. "A Century of International Mortality Trends: II," *Population Studi*es 10(1): 17-42

U.S. Social Security Administration. 2005. "Life Tables for the Social Security Area, 1900-2100." Actuarial Study No. 120 by Felicitie C. Bell and Michael L. Miller (August). https://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl_6_1900.html

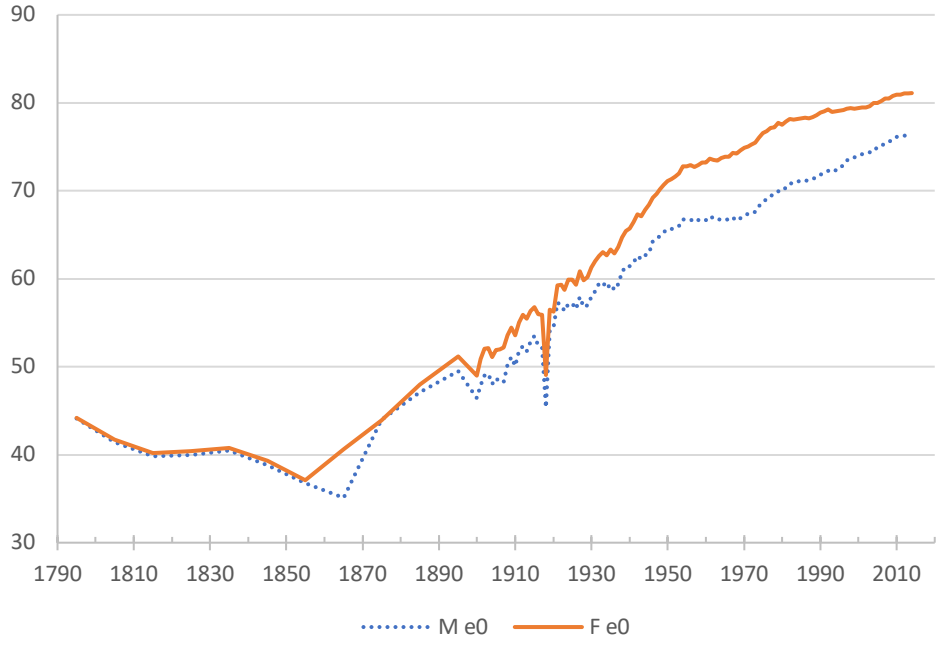
U.S. Social Security Administration. 2017. *Period Life Tables*. On-line data updating Actuarial Study No. 120. https://www.ssa.gov/oact/HistEst/PerLifeTables/2017/PerLifeTables2017.html

Xu, Jiaquan, Sherry L. Murphy, Kenneth D. Kochanek, and Elizabeth Arias. 2016. "Mortality in the United States, 2015," NCHS Data Brief, no. 267. Hyattsville, MD: National Center for Health Statistics.

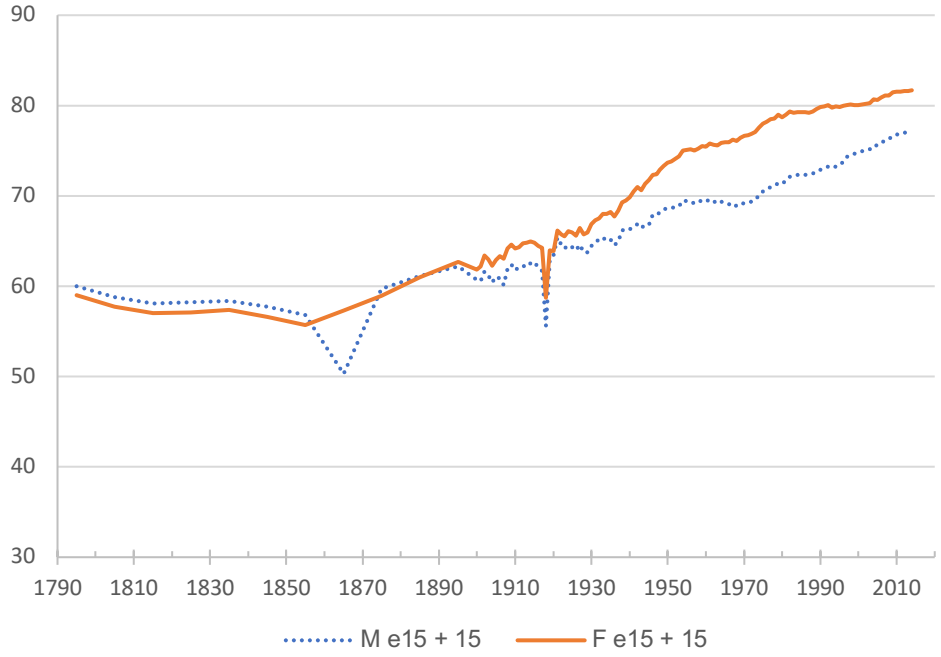
Zeng, Yi, Tu Ping, Gu Baochang, Xu Yi, Li Bohua, and Li Yongpiing. 1993. "Causes and Implications of the Recent Increase in the Reported Sex Ratio at Birth in China," *Population and Development Review* 19(2): 283-302.

Figure 1.1: Life Expectation by Sex at Ages 0 (Birth), 15 and 45: United States, 1795 to 2014

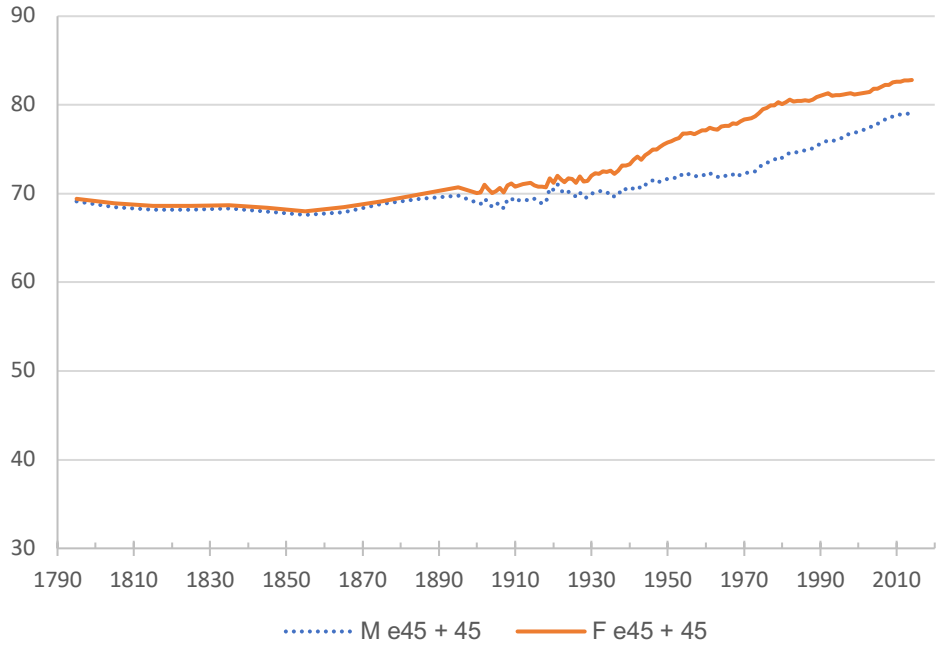
A. Life Expectation by Sex at Age 0 (Birth)



B. Life Expectation by Sex at Age 15



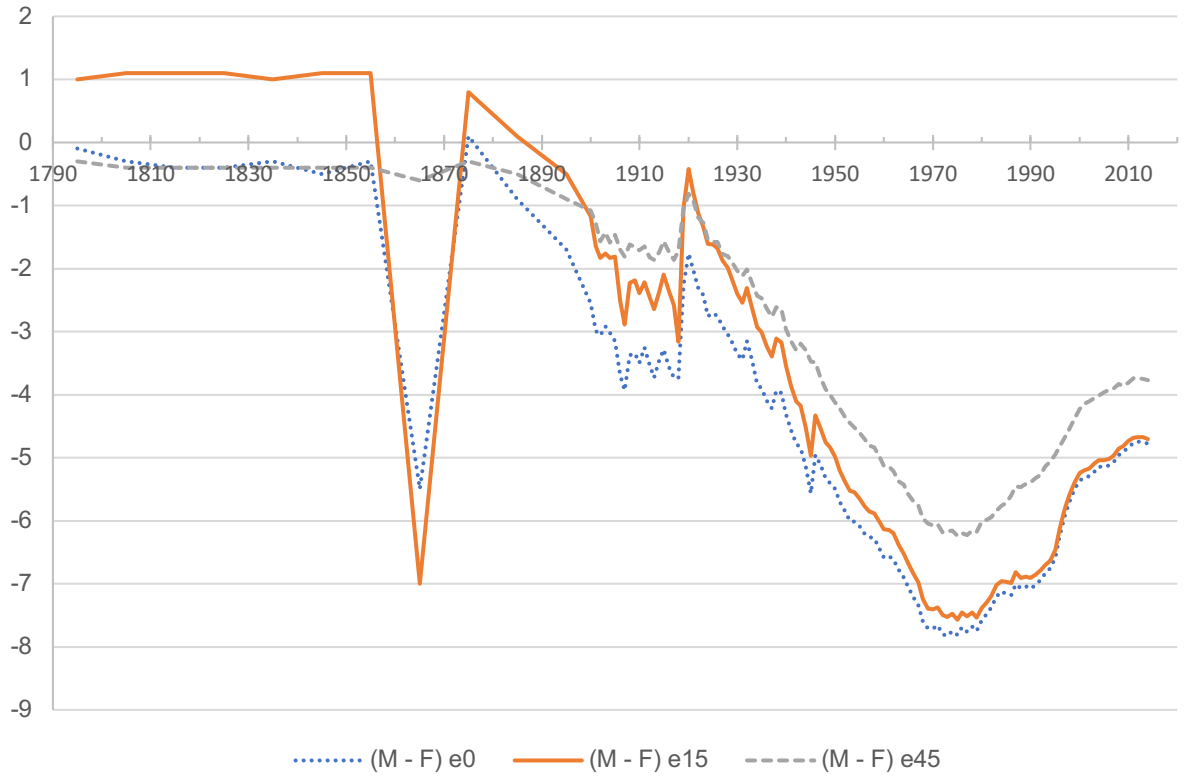
C. Life Expectation by Sex at Age 45



Sources: J.D. Hacker (2010) for the white population; U.S. Social Security Administration (2005, 2017) for the entire population.

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.

Figure 1.2: Male Minus Female Life Expectation: US, 1795 to 2014

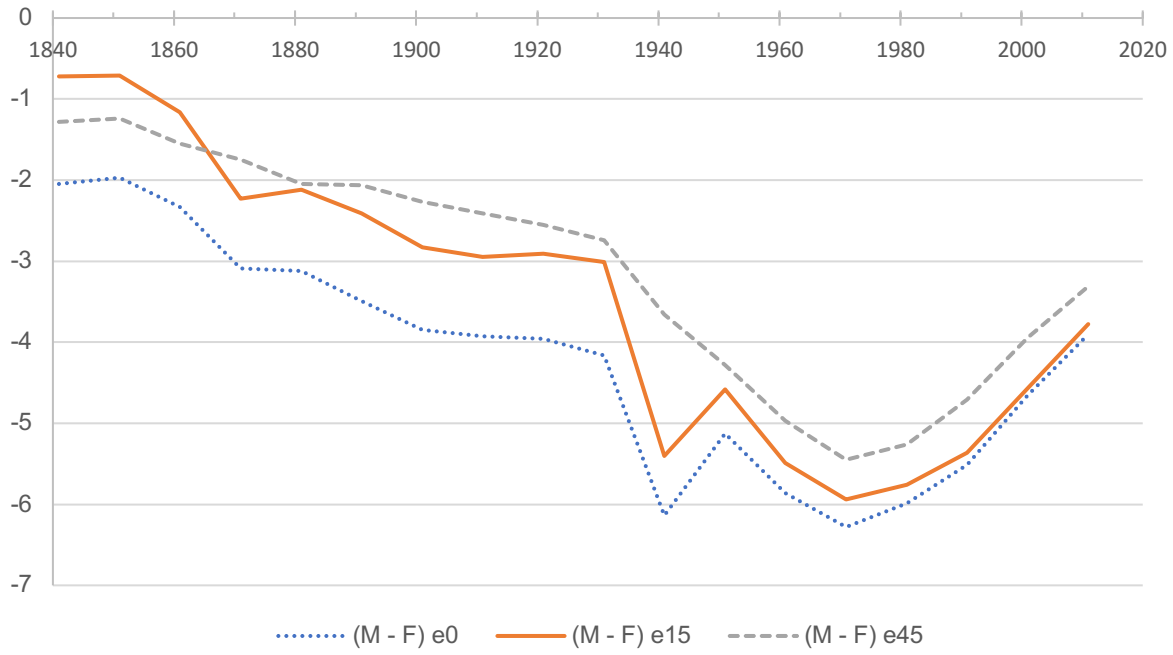


Sources: See Figure 1.1.

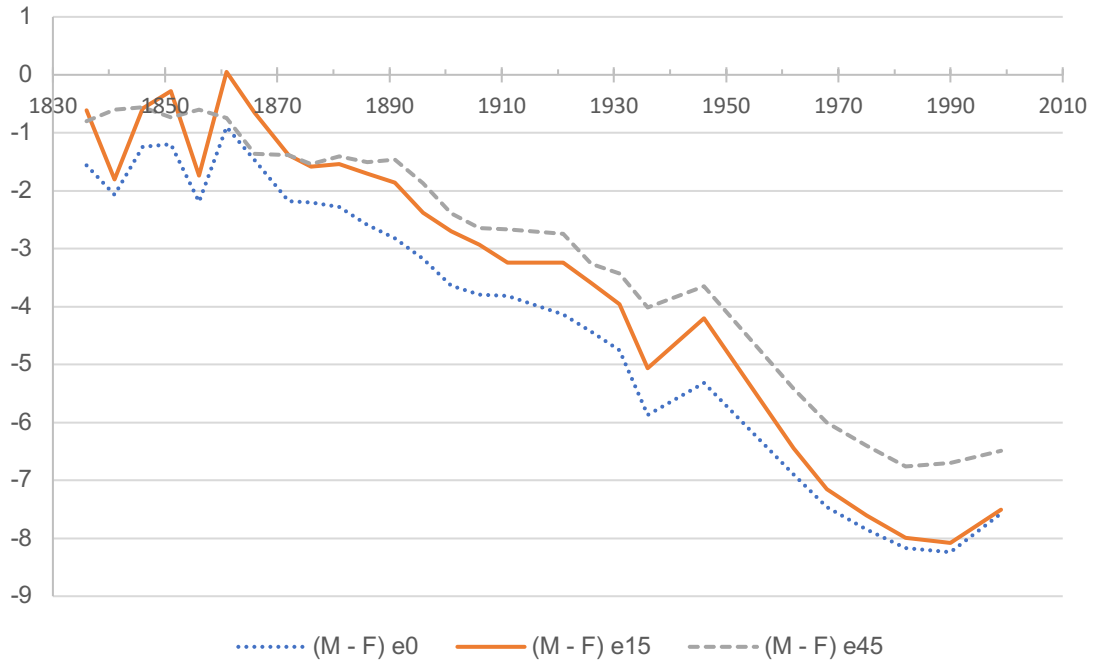
Notes: Shaded region gives the linkage points for the two series: Hacker for whites only and SSA for the entire population. No attempt has been made to link the series. The SSA data are used for 1900 and annual data after to 2015; the Hacker data are used for 1895 (1890-1900) and data on the quinquennial 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).

Figure 1.3: Male Minus Female Life Expectation (Period Rates): England and Wales, France and Sweden, on Census Years

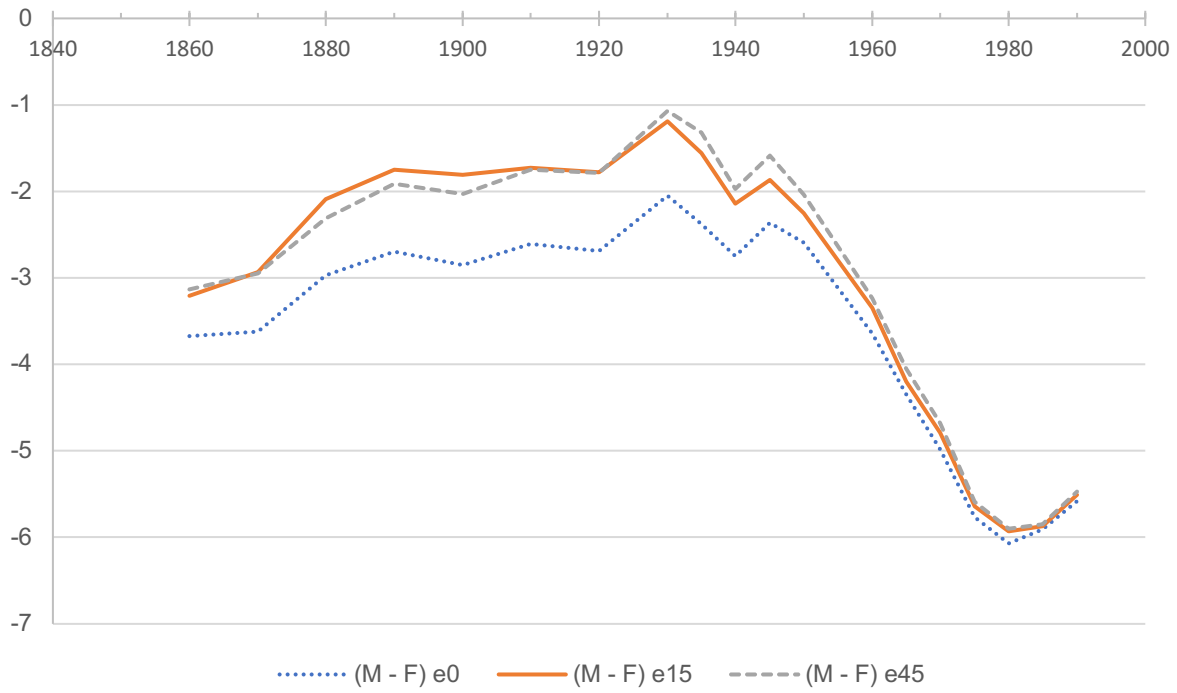
A. England and Wales



B. France



C. Sweden



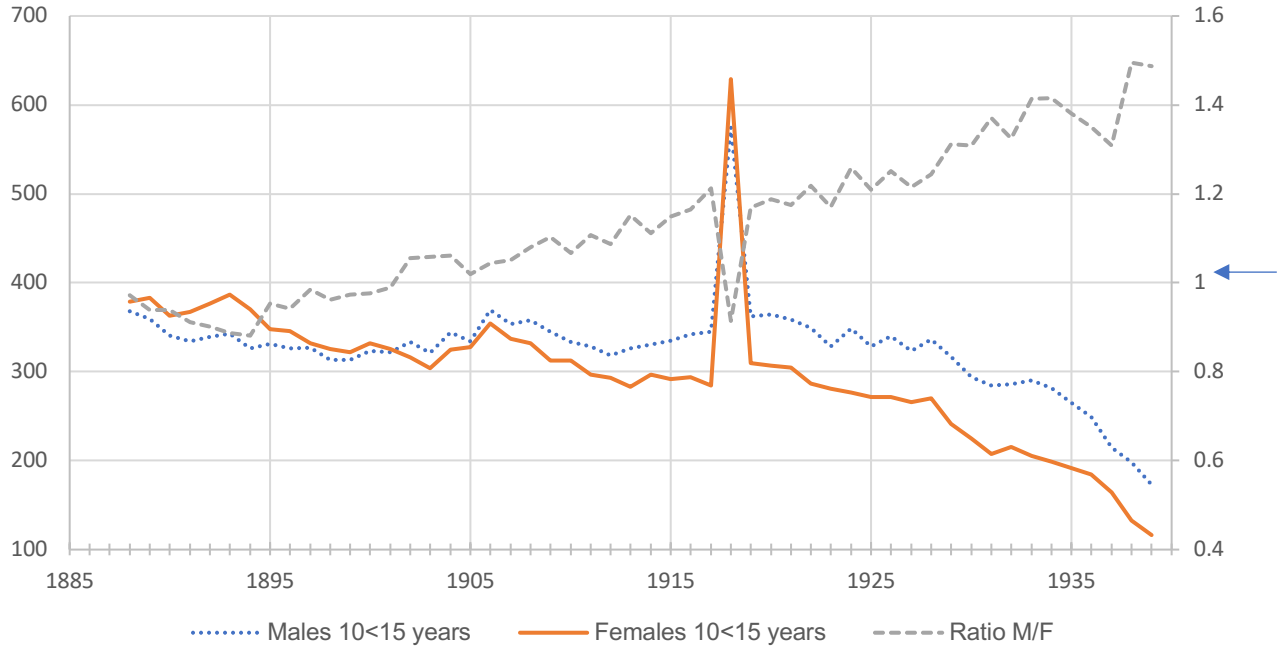
Source: Human Mortality Database, period tables. We only plot years for which censuses were conducted and thus population data by age are accurate. Census data are available every ten years starting in 1841 for England, and every ten years starting in 1860 for Sweden. For France they occured every 5 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.

Figure 2.1: Deaths by Sex and Age: Massachusetts, 1887 to 1940

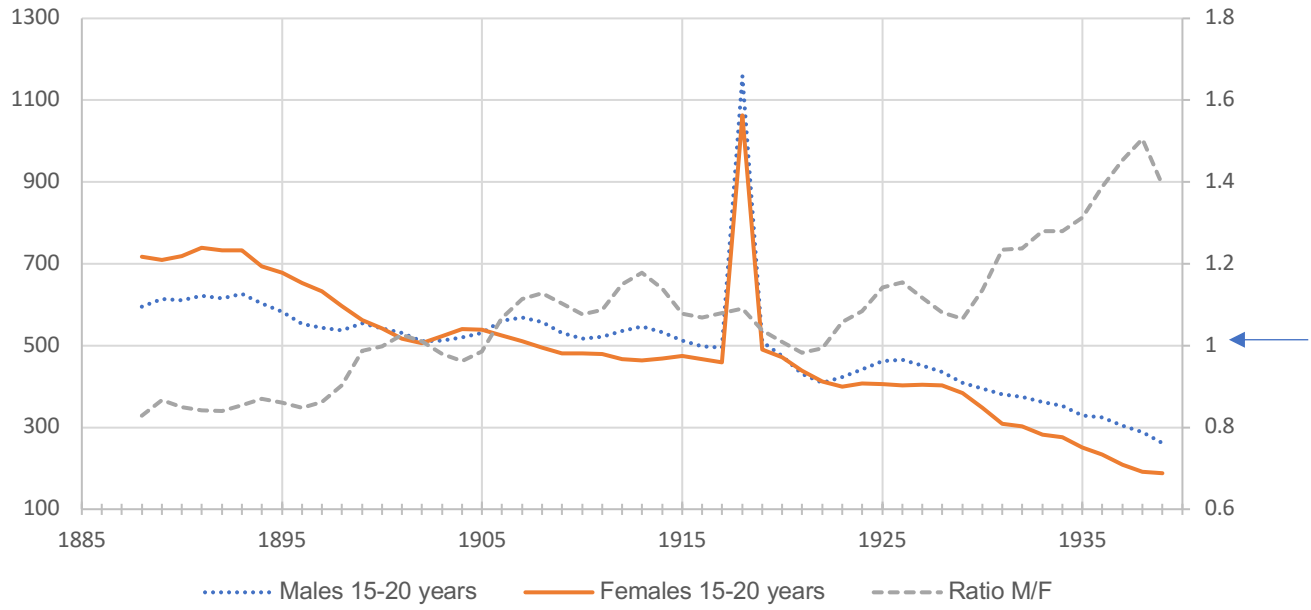
A. Male and Female Deaths, 5 < 10 Years Old and Male to Female Death Ratio (Right Axis)



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)

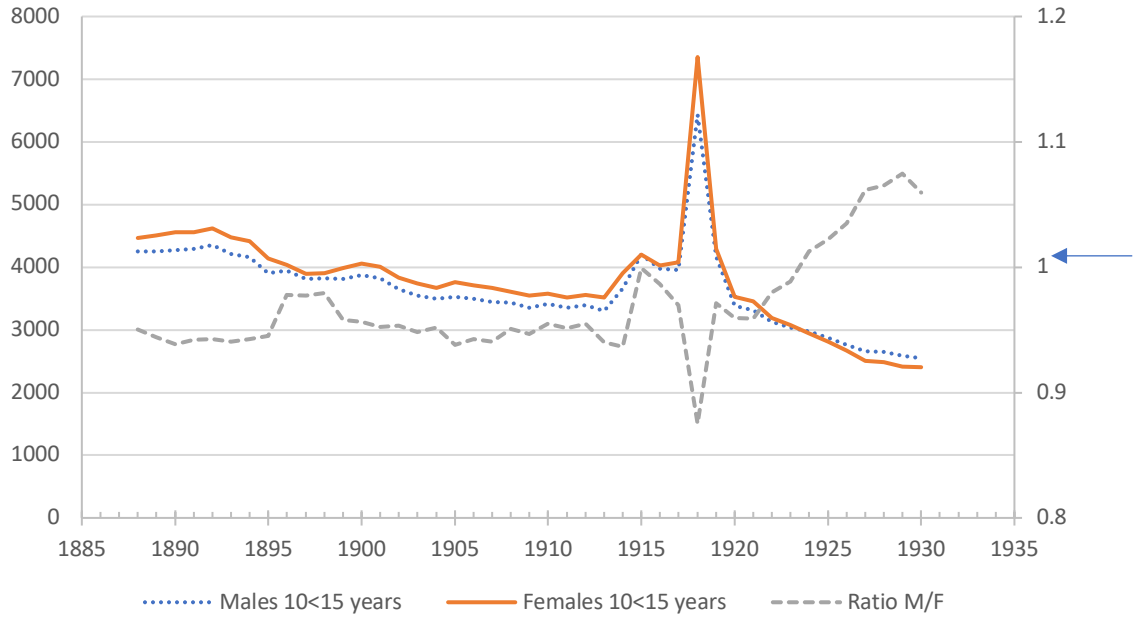


Sources: Commonwealth of Massachusetts (1887 to 1940).

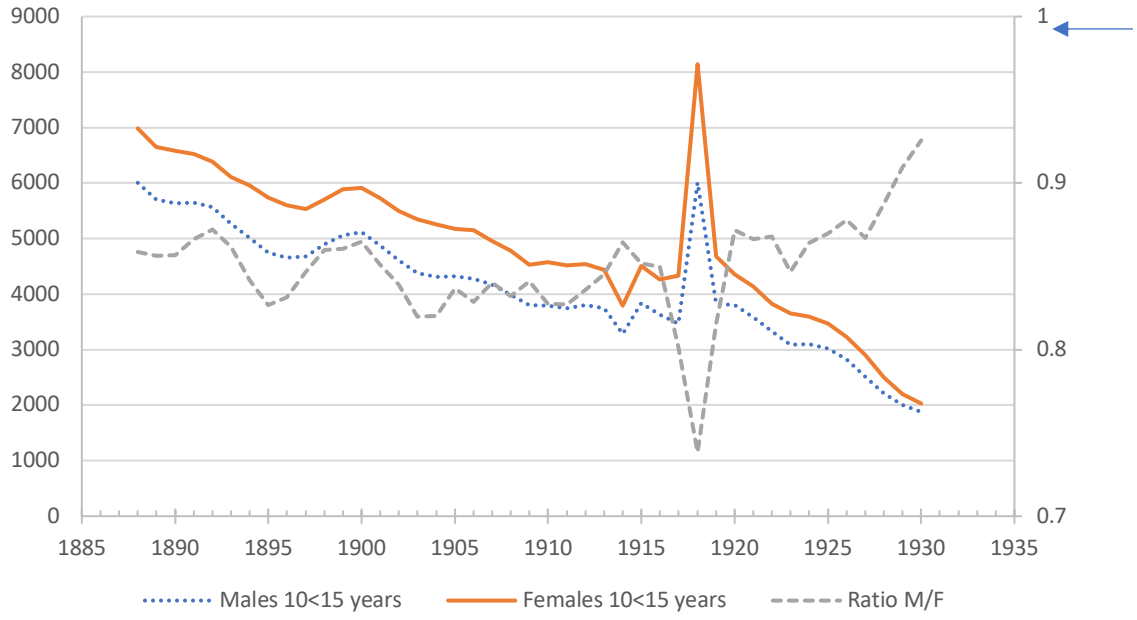
Notes: Data are a three-year moving average of deaths for the age group except that 1918, the year of the influenza pandemic is not part of any average. The two years before and after 1918 are averaged together.

Figure 2.2: Male and Female Deaths, 10<15 Years Old and Male to Female Death Ratio (Right Axis): England and France, 1880s to 1930

A. England

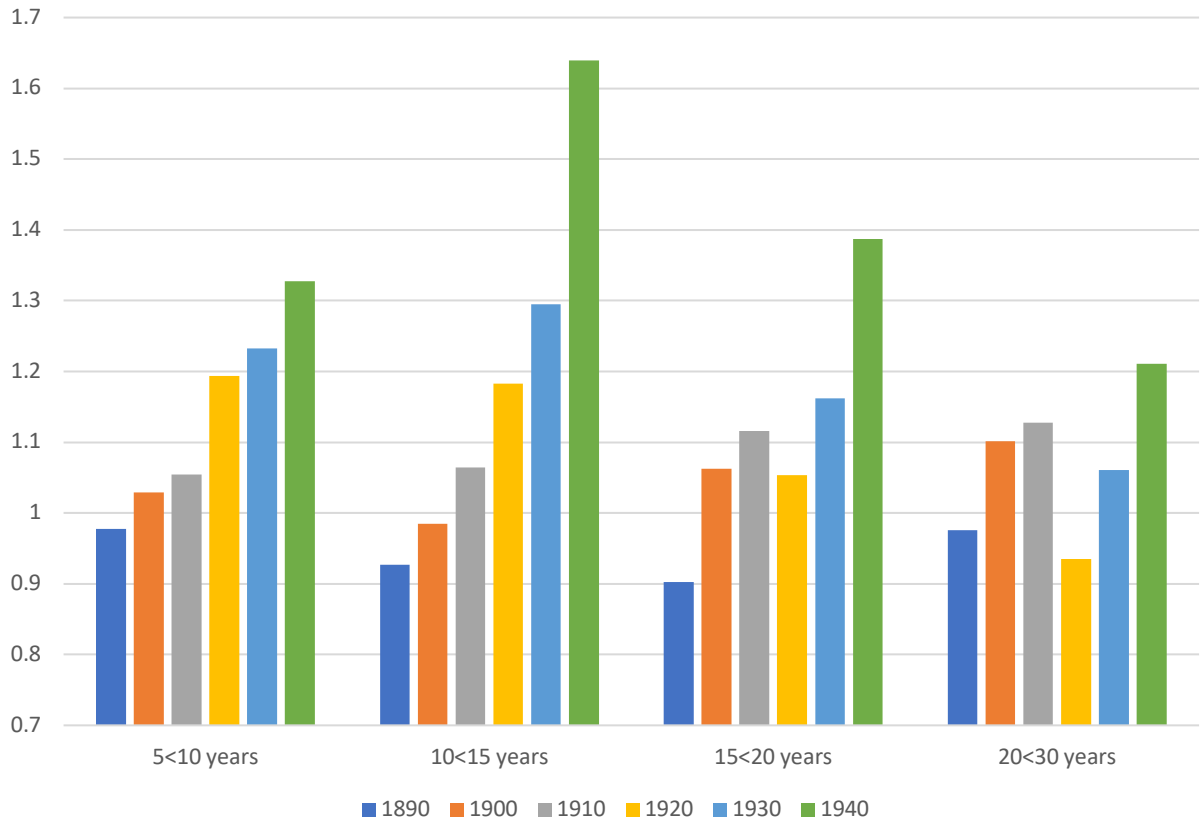


B. France



Source: Human Mortality Database, period tables.

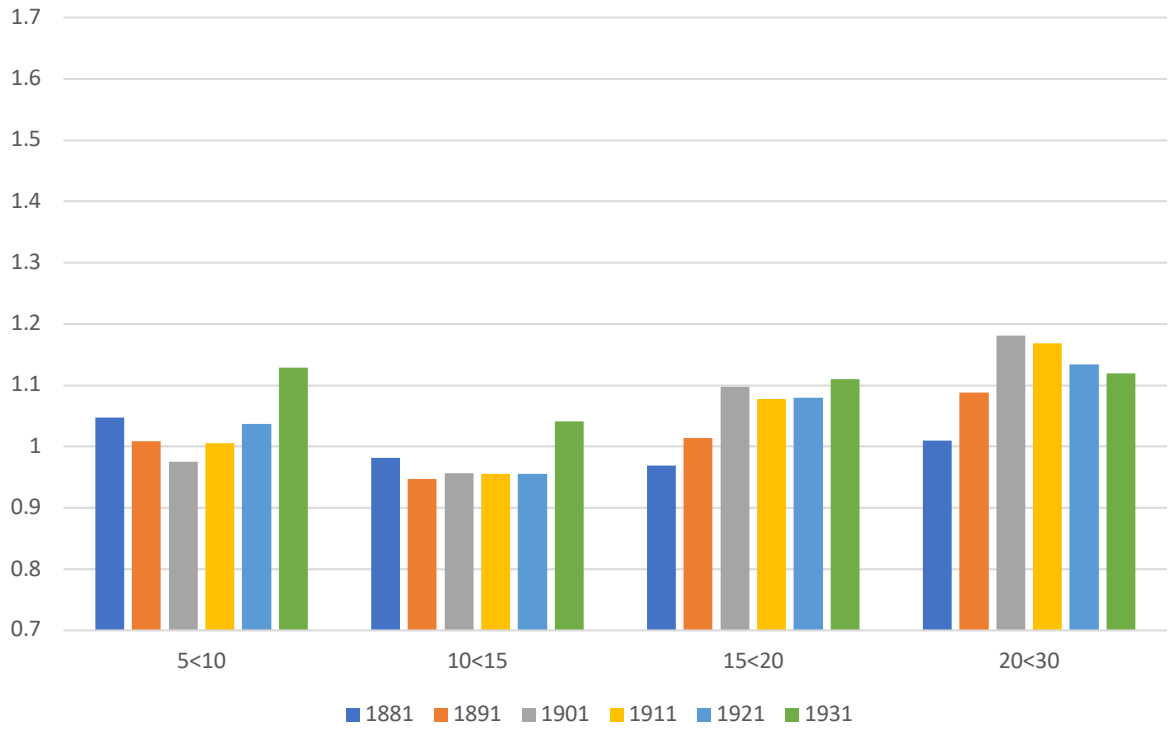
Figure 2.3: Ratio of Male to Female Death Rates for Various Age Groups (5<10 to 20<30): Massachusetts 1890 to 1940



Sources: Commonwealth of Massachusetts (1887 to 1940). U.S. Census of Population, 1890 to 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)

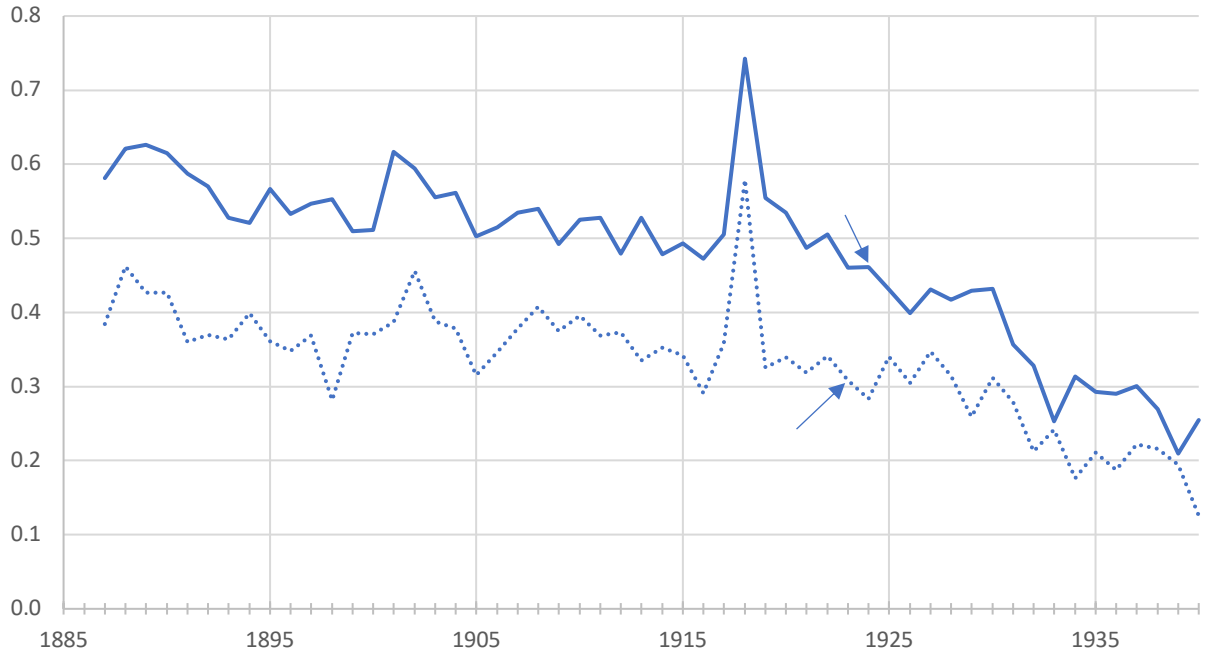
Figure 2.4: Ratio of Male to Female Death Rates for Various Age Groups (5<10 to 20<30), England: 1881 to 1931



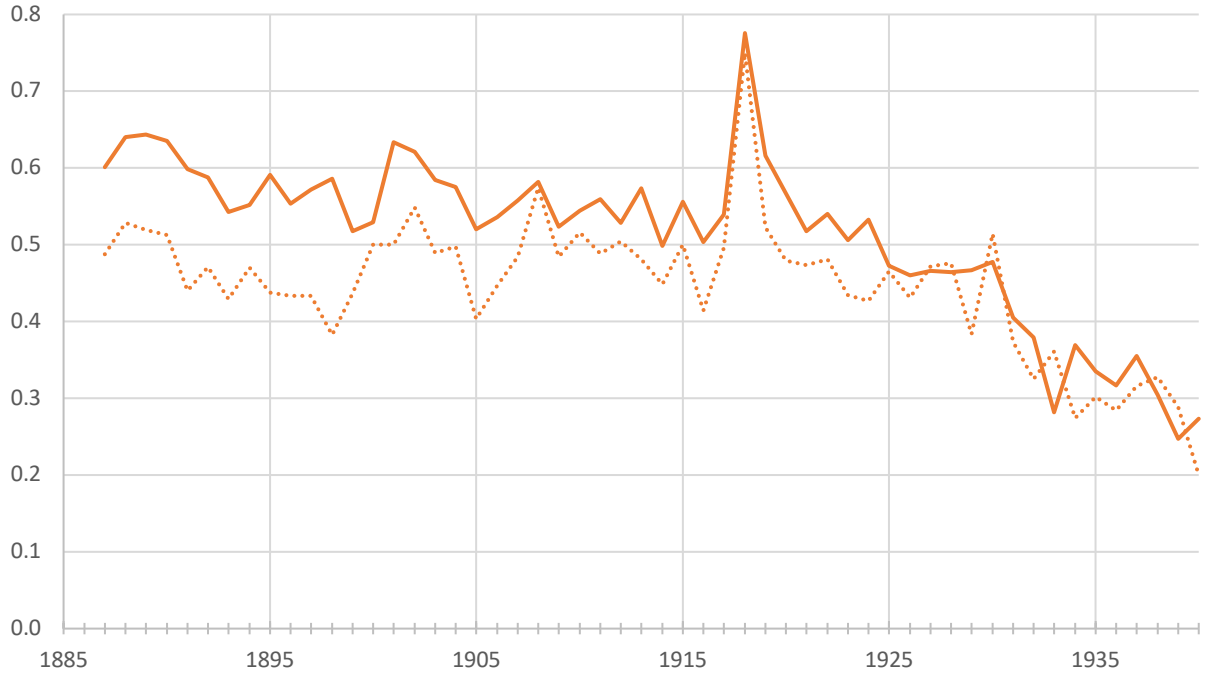
Source: Human Mortality Database.

Figure 2.5: 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



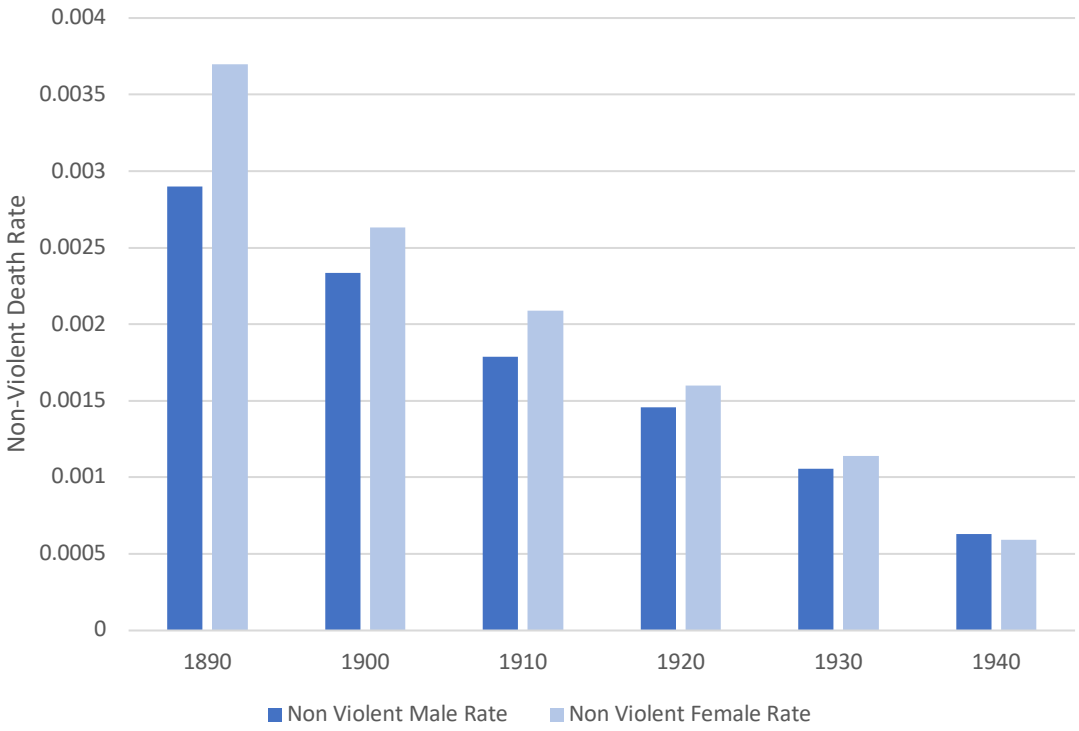
B. Fraction of Non-violent Deaths due to Infectious Diseases



Sources: Commonwealth of Massachusetts (1887 to 1940).

Notes: The blue lines are (infectious diseases/all deaths); the orange lines 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. Shaded areas denote changes in cause of death categories. Actual population values are used rather than three-year moving averages.

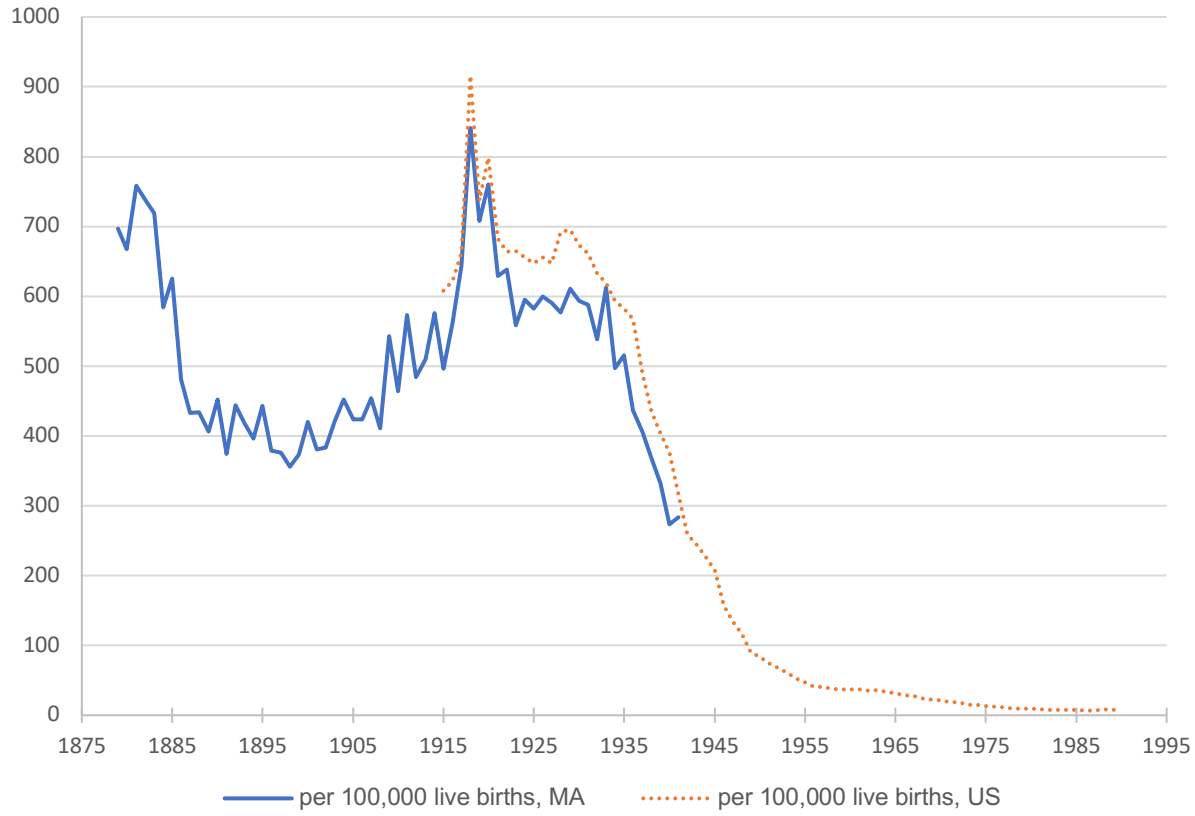
Figure 2.6: Death Rate from Non-Violent Deaths: 10<15 Years Old: Massachusetts 1890 to 1930, by Sex



Sources: Commonwealth of Massachusetts (1887 to 1940); U.S. Census of Population, 1890 to 1940.

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.

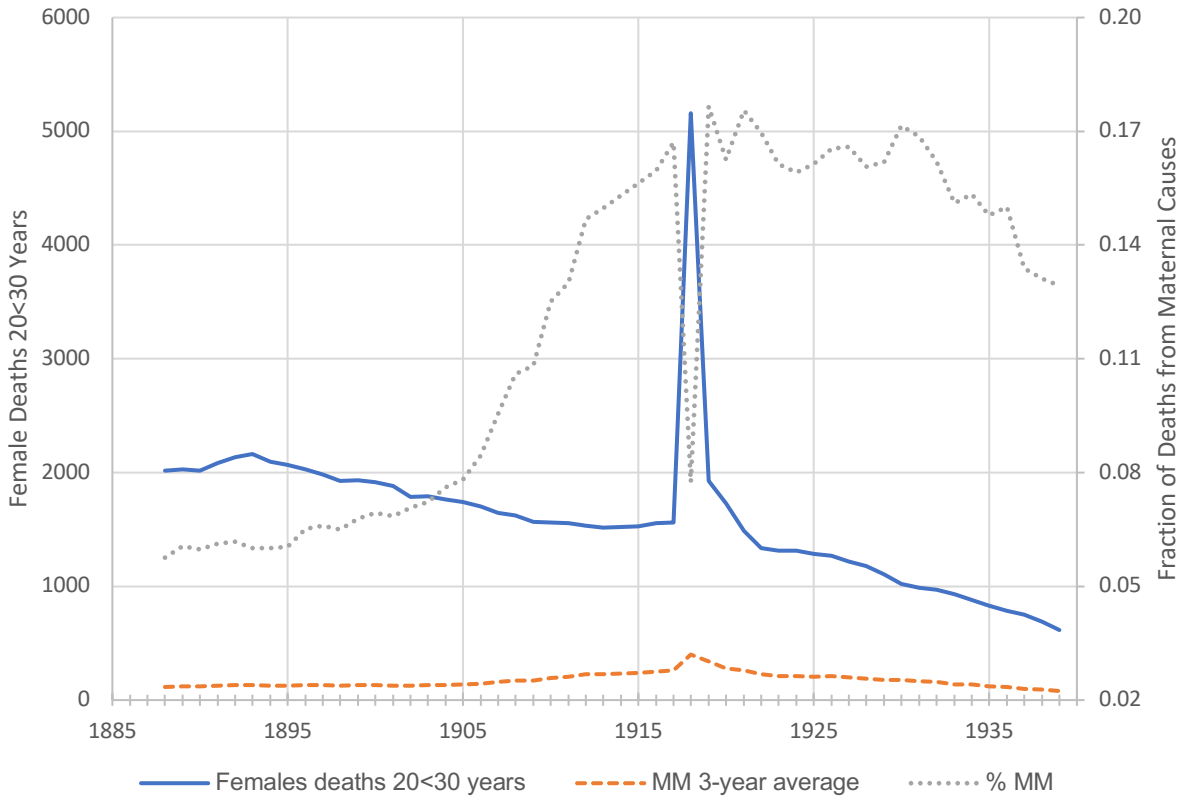
Figure 2.7: Maternal Mortality Rate in Massachusetts: Maternal Deaths/100,000 Live Births, 1887 to 1941



Sources: Commonwealth of Massachusetts (1887 to 1940); *Historical Statistics* (2006), series Ab924.

Notes: For cause of death categories, see notes to Figure 2.8. 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 due to poor original print quality.

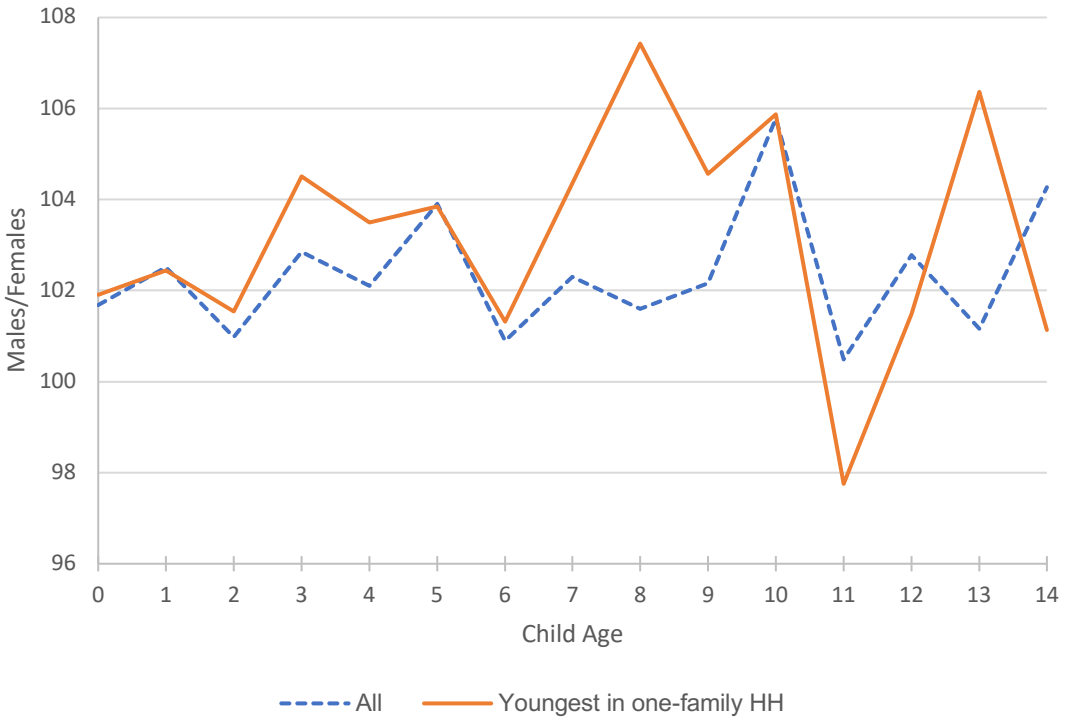
Figure 2.8: Female Deaths, 20<30 Years Old, Deaths due to Maternal Causes, and the Fraction of All Deaths due to Maternal Causes: Massachusetts, 1887 to 1940



Sources: Commonwealth of Massachusetts (1887 to 1940).

Notes: "Female deaths 20<30 years" is a three-year centered moving average, except for 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.

Figure 3.1: Child Sex Ratios (Males/Females) in US Families: 1850 to 1900



Source: 1850, 1860, 1870, 1880 and 1900 Censuses, 1 percent random samples.

Notes: Households containing only one family and with children of only the household head (minors who are adopted or are step-children) are omitted.

Table 3.2: Sex Ratios in the US: 1850 to 1900

	Observed Birth Order of child				
	1	2	3	4+	N
Part A: Sex ratios among Census population in US 1850-1900					
Ages 0 to 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 to 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: 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.

Table 3.3: Heights and Weights of School Children, 5 to 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
Dependent Variable Mean	46.78	3.84	-0.127	50.81	3.91	-0.0554
Female	-0.278***	-0.00595***	0.00323	-1.780***	-0.0362***	0.00350
	(0.0692)	(0.00148)	(0.0317)	(0.212)	(0.00399)	(0.0308)
Irish	-0.460***	-0.00958***	-0.211***	-0.568***	-0.00862**	-0.0746***
	(0.0599)	(0.00128)	(0.0274)	(0.184)	(0.00345)	(0.0266)
Female × Irish	-0.0356	-0.000928	-0.0135	-0.395	-0.00754	-0.0504
	(0.0901)	(0.00193)	(0.0413)	(0.276)	(0.00519)	(0.0400)
Number of observations	9,250	9,250	9,250	9,250	9,250	9,250
R^2	0.700	0.704	0.013	0.589	0.612	0.004

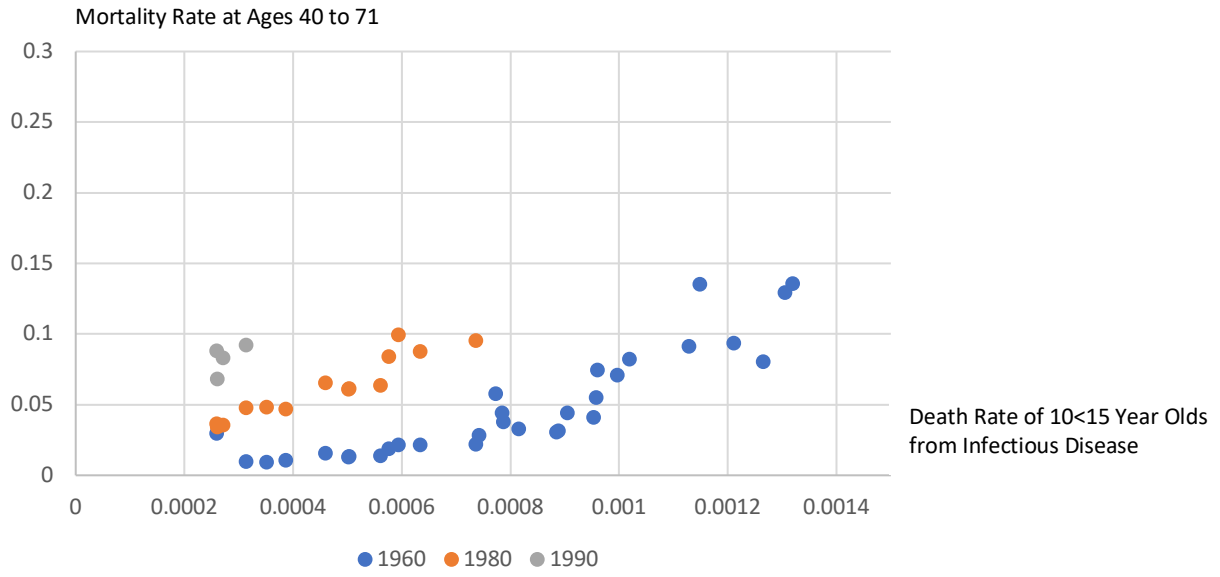
*** p<0.01

Source: Bowditch (1877).

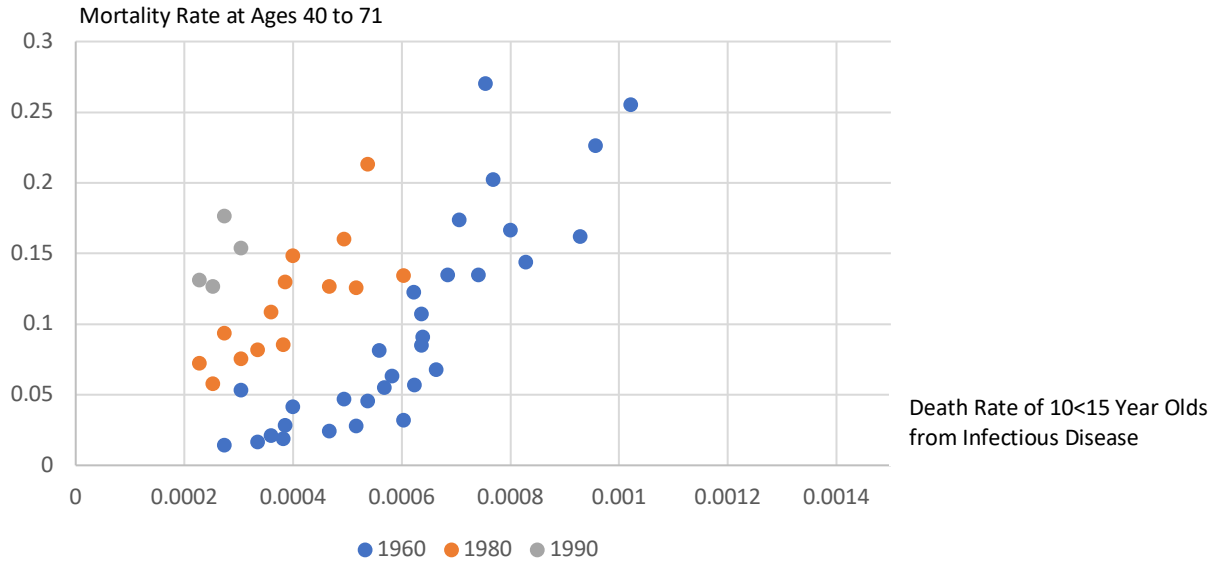
Notes: The Z-score is computed as mean 0 and standard deviation 1 within each sex-age group where the mean and standard deviation come from the native-born parentage group. A full set of age dummies is included in each regression. The weights were given in four-pound bins (e.g., 50-54) and we used the lower bound in these calculations. Note that the lower R^2 for the Z-score regressions is because the age group fixed effects no longer matter.

Figure 4.1: Relationship between the Death Rate from Infectious Disease at Ages 10<15 and the Mortality Rate by Age (to 71 Years) in 1960, 1980, and 1990: Massachusetts

A. Females



B. Males



Sources: Deaths of 10<15 year olds from Commonwealth of Massachusetts (1887 to 1940); annual population of 10<15 year olds estimated from births by year and deaths by age and year. Mortality rate by age from US Vital Statistics.
Notes: Each point is an age in one of three years, where each year includes only ages to 71. The included ages in 1960 is 40 to 71; in 1980 58 to 71 years; in 1990 68 to 71. The X axis is truncated and excludes the death rate of 10<15 year olds during the high infectious disease environment of the flu pandemic of 1918.

Appendix Table 1: Maternal Mortality Calculation

The calculation was done for 1880 to 1990, 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 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):

Age	Life Expectation: Additional Years Left		
	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

Appendix Table 2: 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	1,292***	[39.27]	-34.25	[55.0]	70,541
Property value of mother	44.020***	[4.712]	-0.521	[6.60]	73,399

Sources: 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