US Infant Mortality in the 19th Century: New Evidence from Childhood Sex Ratios

Jesse McDevitt-Irwin and James R. Irwin*

31 October, 2023

Abstract

Basic facts of infant mortality in the 19th-century US are largely unknown, due to a lack of data on births or infant deaths. Contradictory views emerge from previous research. A variety of historical evidence suggests US whites were among the healthiest 19th-century populations. However, according to the latest *Historical Statistics of the United States*, infant mortality among US whites circa 1850–1880 was substantially worse than in much of contemporary Europe. We resolve this puzzle with new evidence on infant mortality: childhood sex ratios. Because of the female survival advantage in infancy, high rates of infant death tend to be reflected in female-skewed childhood sex ratios. We verify the empirical relationship between infant mortality and childhood sex ratios in populations with historical data on both, and demonstrate that sex ratios reveal broad patterns of infant mortality. Turning to the US census for under-five sex ratios, we estimate white infant mortality circa 1850–1880 was around 75 deaths per 1000 – well under one-half the values presented in HSUS, and well below contemporary European levels. By 1900, infant mortality in the US had increased substantially, pointing to the challenges that modernization posed to population health.

*Authors' information: Corresponding author is Jesse McDevitt-Irwin, Columbia University (jrm2275@columbia.edu). James R. Irwin is Professor Emeritus, Central Michigan University. This paper has evolved as part of a broader body of research on historical sex ratios, in process since May 2020. For helpful comments on this paper and previous versions, we thank Douglas Almond, George Alter, Vincent Bagilet, Lou Cain, Joel E. Cohen, Joseph Ferrie, Kris Inwood, Catherine McDevitt, Joel Mokyr, Claire Palandri, Samuel Preston, Jeffrey Shrader, and Rodrigo Soares. We thank participants at the 2021 and 2022 CEA meetings, the Northwestern Economic History workshop, the University of Chicago Demography workshop, and the 2023 PAA meetings for their comments.

Introduction

Infant mortality is a key indicator of population health and living standards, especially historically, when differences in the infant mortality rate (IMR) across populations were far greater than today.[1] Unfortunately, the basic facts of infant mortality in the 19th-century US have yet to be established, because of a lack of records on births and infant deaths.[2] One view emerges from research constructing life tables for the 19th-century US (Haines 1979, 1998; Hacker 2010), which suggest that US whites' infant mortality circa 1850 to 1880 was high by the standards of contemporary Europe. However, a wide range of existing evidence on historical infant mortality in Europe and the US points to the opposite conclusion: that 19th-century US whites had relatively low infant mortality.

In this paper we offer a new empirical basis for characterizing broad patterns of infant mortality, using childhood sex ratios. It is well-known that males are biologically more vulnerable than females to infant mortality; it follows that high rates of infant death tend to be reflected in female-skewed sex ratios among survivors.[3] We document this tendency using historical data assembled from censuses and birth and death registries, and develop a simple empirical model to characterize infant mortality based on childhood sex ratios.

Applying our model to sex-ratio data from the decennial US censuses, we estimate that infant mortality among the white population was in the range of 70 to 80 deaths per thousand live

[1] In the 19th century, across Europe alone, rates ranged from under 100 to over 300. The world has seen a collapse of infant mortality since the early 20th century. By 2020, over one-third of the world population lived in places with infant mortality rates below 10 deaths per thousand births, and two-thirds in places with rates below 30. Authors' tabulations from country data for 2020 reported by the World Bank: Mortality rate, infant and Population, total (both accessed 2022-04-25); Mitchell (1998, pp. 120–122).

[2] The empirical record of infant mortality in much of 19th-century Europe is reasonably complete (see 'Online Appendix 2: Data Sources'), based on records of births and infant deaths which are simply unavailable for most of the US before the early 20th century (Haines 2006; HSUS Ab33,36). Without such records, estimates of infant mortality for the 19th-century US as a whole are conjectures from model life tables. Most prominent are those of Haines (1979, 1998) and Hacker (2010); note that those studies seek to characterize 19th-century US mortality across the lifespan, and are not focused on estimating infant mortality. Hacker (2010, p. 76) explicitly points out the need for further research on infant and childhood mortality during the period.

[3] Of course, this result will not hold if extremes of sex and gender discrimination reverse females' biological survival advantage, as seen in cases of "missing women" *à la* Sen (1992).

births in the period 1846–1880. Our findings sharply contradict the IMR series (Ab921) presented in the most recent edition of *Historical Statistics of the United States* (henceforth, HSUS), where infant mortality rates for whites range from 176 to 217 in the period 1850 to 1880. According to the HSUS series, US whites suffered rates of infant mortality worse than those of contemporary England or France. On our evidence, US whites were among the healthiest populations of the 19th century.

Historical Background

The basic facts of infant mortality for US whites since the mid-19th century may appear to be reasonably complete.[4] The most recent (2006) version of *Historical Statistics of the United States* (HSUS) presents the infant mortality rate (IMR) for the white population at decennial benchmarks from 1850 to 1910, and annually starting in 1915.[5] Figure 1 plots this series against the backdrop of IMR data available for a cross-section of European populations from 1840–1990. The HSUS series features high rates of infant mortality across the census benchmarks from 1850 to 1880, averaging just under 200, before dropping fairly steadily to just under 100 by 1910. After something of a pause until 1918 (and the great flu pandemic), the decline in the series resumes, falling well below 30 points by 1950, and below 10 points by the early 1980s. Looking at Figure 1, the HSUS series falls well within the range of European infant mortality experiences; arguably, what stands out is a general pattern of massive improvement in infant mortality since the late 19th century.

The substantial decline in the HSUS infant mortality series after 1880 might seem to be just

[4]The empirical record of Nonwhite and Black infant mortality is clearly incomplete (HSUS Series Ab922, Ab923), and largely outside the scope of this paper.

[5]HSUS Series Ab921. The annual IMR series, from 1915 on, was presented in previous editions of HSUS (1949, 1952, 1960, 1976). The most recent (2006) edition added the census benchmark values of IMR for 1850 to 1910, from Haines (1998). See below for further discussion of the HSUS series, but note that the value presented for the year 1910 is an estimate for circa 1904, based on 1910 census data (Haines 1998, p. 154,167; HSUS 2006. Table Ab1-10, footnote 2). Note also that prior editions of HSUS were produced and published by the US Bureau of the Census; the current (2006) edition was "prepared by the academic community", with Michael R. Haines the editor of "Chapter Ab – Vital Statistics" (HSUS 2006: Appendix 3, "Editions and Copyright" and "Editor's Preface"; Haines 2006).

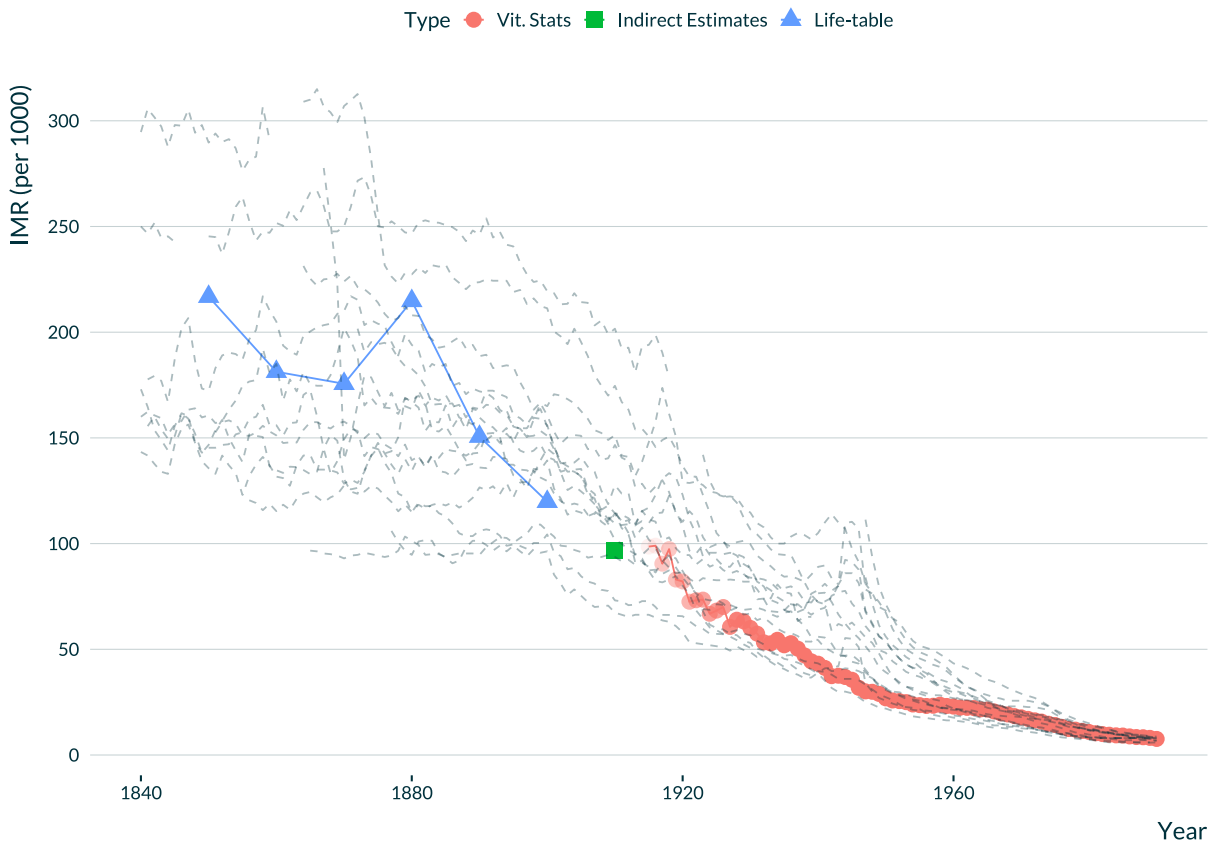

Figure 1: Infant Mortality Rates, 1840-1990. The colored points represent US whites; the dashed lines are 3-year rolling averages of various European populations. Sources: HSUS series Ab921 and see data appendix. US white Infant Mortality by Year, from HSUS, colored by the type of estimate. 'Life-table' refers to estimates that are extrapolations from older age mortality using life tables (Haines 1979, 1998). The indirect estimate is from Haines (1998, table AI) which used the surviving-children method with data from the 1910 census (maternal recall and population by age), building from Haines and Preston (1997). The value presented for 1910 is actually an estimate for circa 1904, but we have plotted the values as reported in the HSUS series; further to this point, see Figure 5 and related text (below). The vital statistics series relies on data from birth and infant death records in the US Birth Registration Area (BRA); these values are shaded according to the proportion of US population covered by the BRA, going from 1/3 in 1915 to 100% in 1933 (HSUS Series Ab33).

another facet of the widely studied "mortality transition" (Caldwell 2001), which has seen life expectancies soar and mortality rates plummet across the globe since the early 20th century.[6] The mortality transition forms a dominant paradigm for historical demographic research, framing a wealth of studies which investigate the emergence of the very low mortality regimes enjoyed in the 'developed' world.[7] Central to this paradigm is a presumed historical fact – that rates of infant and child mortality were bitterly high in the pre-industrial and early-industrial past. This widely-shared understanding of history is evident in the opening line of Anderson, Charles, and Rees (2022, p. 126): "Since the mid-19th century, mortality rates in the Western world have plummeted and life expectancy has risen dramatically." Central also to the paradigm is a broad historical explanation for the collapse of mortality: that the scientific and industrial revolutions were transmuted into mortality decline (Caldwell 2006: chapter 8). The collapse in mortality thus plays a central role in broader narratives of progress surrounding 19th and 20th century modernization and industrialization. In this context, the tremendous decline of infant mortality seen in the HSUS series may seem unremarkable.[8]

However, the levels of infant mortality presented in HSUS for the 19th-century US are not credible in light of infant mortality evidence from both sides of the Atlantic. Infant mortality around 200 would rank US whites among the unhealthier contemporary European populations, with double the infant mortality of Norway (IMR of 100), much worse than England or France (IMR 150–160), and not much better than Germany and Austria (IMR

[6]Global average life expectancy at birth has risen from 32 in 1900 (Riley 2005b, table 1) to 73 in 2019 (UN Population Division 2022), and over a similar period under-5 mortality has plummeted from roughly 1-in-2 (Hill 1995) to under 1-in-25 (UNICEF Data, accessed 2023-10-31). See also Riley (2005a) and Costa (2015), who use the term "health transition" to refer to the widespread huge improvements in human health.

[7]Among many contributions addressing the mortality transition, see, for Europe: Preston and Van de Walle (1978); Woods, Watterson, and Woodward (1988); Haines and Kintner (2000); Kesztenbaum and Rosenthal (2017); and for the US: Condran and Cheney (1982); Haines (2001); Cutler and Miller (2005); and Alsan and Goldin (2019).

[8]This view is reinforced by Hacker's (2010) life tables for the 19th-century US white population. The IMRs from these life tables also average about 200 for the period 1850-1870, and Hacker's life-table IMR estimates exhibit a strong downward trend from the 1860s, fitting the mortality transition paradigm even better than does the HSUS series.

above 200). Such poor health is difficult to reconcile with the substantial evidence that US whites were among the tallest people in the world during the 19th century (Fogel et al. 1983, p. 463).[9] In addition, the US was predominantly rural during the 19th century, when there was a substantial "urban penalty" (Kearns 1988) in mortality.[10] In the 19th century, US whites were a tall, mainly rural population drawing millions of voluntary immigrants from Europe: infant mortality of 200 deaths per 1000 would constitute a serious puzzle.

Attention to the slim body of direct evidence on infant or childhood mortality in the 19th-century US also casts doubt on the high level of infant mortality shown in the HSUS series. Most notably, there are credible data on births and infant death for the state of Massachusetts from circa 1860 (Abbott 1897, p. 714; Haines 2006, p. 385). There we see infant mortality averaging less than 160 for the period 1860–1880 – a remarkable 40 points below the level of infant mortality in the HSUS series for US whites. But births and infant deaths records from 1890 and 1900, covering many more states, clearly identify Massachusetts as a high mortality state, as expected from its relatively high level of urbanization.[11] In sum, it is simply not credible that infant mortality among 19th-century US whites exceeded that of the State of Massachusetts.

All of this evidence points to Massachusetts having greater infant mortality than the US as a whole during the 19th century.

Furthermore, the limited evidence on child survival from the rest of the US suggests that

[9]US whites were one of the tallest populations in the world throughout the 19th century. Native-born US white men stood at over 170 cm in mid century, while the French were around 165 (Weir 1997, table 5B.1) and Austrians and Italians even shorter (Komlos and Baur 2004, table 1). In healthier European countries (lower IMR), like Scotland and Norway, men were taller, approaching 170 centimeters, though still shorter than US whites (Komlos and Baur 2004, table 1). Sample-selection issues complicate an uncritical interpretation of 19th-century US height data (for discussion, see Bodenhorn, Guinnane, and Mroz 2017), but would be unlikely to reverse the US height advantage.

[10]For example, in 1890 England, urban infant mortality was about 220, while rural was just under 100 (Woods, Watterson, and Woodward 1988, p. 353). On the urban mortality penalty in the 19th century, among many possible, see also Davis (1973, pp. 102–104), Williamson (1982), Haines (2001), and Cain and Hong (2009).

[11]For example, in 1900, when more vital-statistics data are available, the infant death rate in Massachusetts (86% urban) was 182, while in Michigan (40% urban) it was 128 (authors' calculations, based on data in Condran and Crimmins 1980, table 1). More generally, Massachusetts' infant mortality appears to have been typical of the highly urbanized Northeast region (Condran and Crimmins 1980).

infant mortality was more likely 100 than 200 in the mid-19th century. Lynch, Mineau, and Anderton (1985, table 4) find that infant mortality in Utah from 1850–1880 was around 100, just half the level found in the HSUS national series. Haines's (1977) results for upstate New York in 1865 point in a similar direction. Using census data on maternal recall, Haines (1977, table 4) estimates rural under-5 mortality of 18–19% and urban 25–26% (table 4). Those levels of child mortality might suggest rates of rural and urban infant mortality around 110 and 150 respectively, based on the oft-cited Coale-Demeny West model life tables.[12]

More broadly, Preston and Haines (1991, p. 53–57) present a variety of infant mortality estimates for US cities circa 1850–1880, with rates around 165–175 for Philadelphia, 180 for Brooklyn and Chicago, and 170–200 for Boston. In light of those city-level estimates, the HSUS values are implausibly high – around 200 deaths per 1000 across the period 1850–1880 – for a population that was nearly three-quarters rural in 1880.
Indeed, death data from around the turn of the century clearly show that the highly urbanized Northeast region had higher infant mortality than the rest of the country.[13]

In sum, given known patterns of population health in both Europe and the US, the high levels of infant mortality presented in the HSUS for 19th-century whites are puzzling. But the 'puzzle' has an easy resolution – the HSUS IMR values for 1850 to 1900 are conjectural. Lacking evidence on births or infant deaths during this period, these 19th-century infant mortality rates are extrapolations from mortality at older ages. In sharp contrast, for the 20th century, the HSUS series reflects direct and indirect estimates of infant mortality.

In terms of sources and methods, the HSUS series (Ab921) includes three different types of estimates. First, the annual values (1915 on) are direct estimates of infant mortality

[12]We take these values from the Coale-Demeny West model life tables, the average of levels 10 and 11 for urban, and 14 and 15 for rural (Coale and Demeny 1983, pp. 46–49). Similar rates of infant mortality are consistent with the child mortality rates found by Hacker et al (2023, table 2) for the US in the 1850s.

[13]See Condran and Crimmins (1979, 1980) for 1890 and 1900 infant death rates by state. The 1900 Death Registration Area (DRA) data show much higher rates of infant death in states of the Northeast (35–38) than those of the Midwest (23–25) (authors' calculations).

from registration of births and infant deaths.[14] Second, the value for 1910 is a standard indirect estimate of infant mortality, using census data on maternal recall of children-born and surviving.[15] Third, the decennial benchmark values from 1850 to 1900 are from model life tables, with no basis in data on births and infant deaths.[16] Lacking requisite data for direct or indirect estimates of infant mortality, Haines (1979; 1998) fitted model life-tables to census mortality data for ages 5 to 20.[17] The estimated life tables include the level of infant mortality for each census year (1850 to 1900), which appear in HSUS Series Ab921.[18]

However, infant mortality has no necessary relationship with mortality at older ages.[19] As emphasized by Woods (1993, p. 217), indices of mortality in infancy, early childhood, and adulthood are all "indispensable" for characterizing a population's mortality, because "each one captures a distinctive aspect of the mortality pattern and their empirical interrelations clearly were not predictable in the past." For example, in England from 1840–1880 age 5–20 mortality declined by half, while infant mortality was roughly constant. Here, extrapolating from age 5–20 mortality would produce severe overestimates of past infant mortality. More generally, the highly credible life tables of the Human Mortality Database (HMD) show a wide range of infant mortality rates for given levels of mortality at older ages.[20] Figure 2

[14]Nationwide birth and infant death records start in 1933. From 1915–1932, the estimates cover just part of the country: the 'Birth Registration Area' (BRA), covering about 1/3 of the US population in 1915, increasing to 95 percent coverage in 1932 (HSUS series Ab33).

[15]Though labelled 1910, the estimate is for circa 1904, based on maternal recall in the 1910 census (Preston and Haines 1991, p. 74; Haines 1998, p. 154; HSUS 2006: Table Ab1-10, Footnote 2).

[16]HSUS series Ab9 also presents the decennial benchmark values of white infant mortality, with the following footnote: "For the expectation of life at birth and the infant mortality rate, the values for 1900 and 1910 are from approximately 1895 and 1904, respectively" ((HSUS 2006: Table Ab1-10, Footnote 2). The footnote is correct for 1910 (see note above), but not for 1900. Although Haines (1998, p. 154, 165) includes an indirect estimate of white infant mortality for circa 1894–1895, based on maternal recall in the 1900 census, the HSUS IMR series has the model life table value for 1900, from Haines (1998, p. 160; 1979, p. 307). On the indirect estimates, see also Haines and Preston (1997).

[17]Haines restricts his data to ages 5–20 because these census mortality data are judged to be more complete than the data for other ages (Haines 1977).

[18]Hacker (2010) similarly estimates life tables for the 19th-century US white population, but based on existing estimates of life expectancy at age 20 rather than census mortality data. The associated infant mortality values are broadly comparable to Haines (1979).

[19]Hacker (2010, table 6) makes this point within the context of the 19th-century US, illustrating the wide range of infant mortality possible when extrapolating from adult mortality.

[20]HMD life tables are for "populations where death registration and census data are virtually complete" (HMD "Scopes and basic principles", accessed October 27, 2023).

plots infant mortality rates against age 5–20 mortality rates from HMD life tables, covering a range of European (or European-descent) populations in the period 1835–1925; the shaded area shows the range of age 5–20 mortality rates in Haines's life tables that produced the HSUS IMR estimates (1998, appendix A). With this range of age 5–20 mortality, infant mortality rates in the historical life tables ranged from below 70 to above 200.

Over a decade ago, Hacker (2010, p. 76) noted the problems associated with estimating infant mortality from mortality at older ages, concluding that "empirical research on infant and child mortality in the United States is sorely needed." So far, a lack of birth and infant death records has stood in the way of such research. Here we offer a new empirical approach. Building off well-known facts of biology and demography, we have devised a new method for characterizing patterns of infant mortality, using readily available census data on childhood sex ratios. Relative to values presented in the current HSUS, our evidence points to dramatically lower infant mortality rates among 19th-century US whites, much as one would expect given known patterns of population health in both Europe and the US at the time.

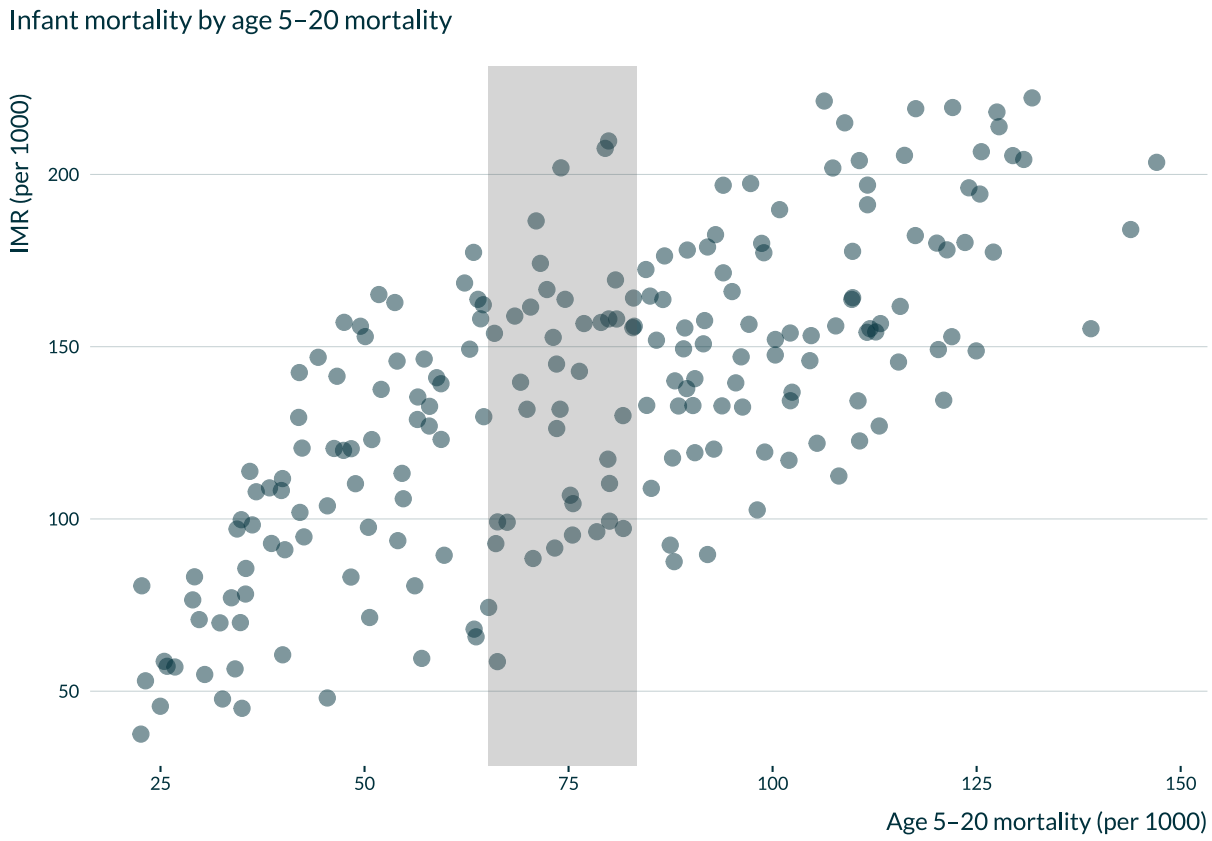

Figure 2: Infant mortality by age 5–20 mortality. Source: HMD life tables 1835–1925. The shaded area is the range of estimates of age 5–20 mortality for US whites 1850–1880, as given by Haines (1998, p. 156–165).

Infant Mortality and Childhood Sex Ratios

It has long been known that biologically, girls are less vulnerable than boys to infant mortality.[21] The corollary which we highlight, and build from, is that high rates of infant mortality tend to skew childhood sex ratios toward females.[22] This pattern is apparent both in historical populations and familiar model life tables. For example, in 1900 infant mortality in Austria was above 200, and there were similar numbers of boys and girls under the age of five. By 1970, infant mortality had plummeted to 20 deaths per 1000 and there were 5% more boys than girls, a value typical of the sex ratio at birth in healthy populations (Maconochie and Roman 1997; Grech, Savona-Ventura, and Vassallo-Agius 2002). A similar pattern is found in a wide range of polities (see below, Figure 3). In familiar model life tables, the pattern is also present: for example, in the Coale-Demeny West model, moving from level 9 to level 22, infant mortality plummets from 193 to 27 (per 1000) and the sex ratio among survivors to age five ($_5l_0$) shifts 2.7 percentage points away from girls (Coale and Demeny 1983, p. 47,52).

A simple model illustrates the impact of infant mortality on the sex ratio among surviving children. Consider the childhood sex ratio as the (natural logarithm) sex ratio of a hypothetical population of survivors to age 1, $ln(\frac{l_1^f}{l_1^m})$.[23] With B^j the number of births and q_0^j the

[21] Current knowledge is conveniently summarized by the editors of PLOS Medicine in their summary of Sawyer (2012): "Newborn girls survive better than newborn boys because they are less vulnerable to birth complications and infections and have fewer inherited abnormalities. Thus, the ratio of infant mortality among boys to infant mortality among girls is greater than one, provided both sexes have equal access to food and medical care." Knowledge of excess male infant mortality dates back at least to the 18th century, for example, Struyck (1740), Wargentin (1755) and Clarke (1786); for discussion, see Théré and Rohrbasser (2006). The female survival advantage in infancy is attributed to multiple factors: females have fewer congenital diseases owing to their redundant X chromosome, and they are also more resistant to infectious disease. For a review see Waldron (1998, p. 64–83).

[22] Our method would not apply to societies with 'missing women': i.e., where the "social vulnerability" of girls outweighs the "biological vulnerability" of infant males (Thompson 2021, p. 467). There, male-skewed sex ratios in the context of high infant mortality are a clear sign of 'missing women'. See, for example, Beltrán Tapia and Raftakis (2021), especially Figures 2 & 3. Existing evidence on child mortality from the 19th-century US shows a clear female survival advantage (Haines 1977, table 7; Kunze 1979, table 14; Lynch, Mineau, and Anderton 1985, table 4), so 'missing girls' are not a concern for our findings.

[23] We adapt life-table notation from Preston, Heuveline, and Guillot (2001, chapter 3). Note that observed sex ratios are for populations in an age interval ($_nL_x$) but we model populations surviving to exact age 1 (l_x), a simplification that clarifies the key factors determining childhood sex ratios. Mortality among ages

infant mortality rate of sex j; we can express the sex ratio as follows:

$$ln(\frac{l_1^f}{l_1^m}) = ln(\frac{B^f \cdot (1 - q_0^f)}{B^m \cdot (1 - q_0^m)}).$$

A few steps of algebra give us the following expression:

$$(1) \quad ln(\frac{l_1^f}{l_1^m}) = ln(\frac{B^f}{B^m}) + [ln(1 - q_0^f) - ln(1 - q_0^m)].$$

Here we see that the sex ratio at age 1 is determined by two additively separable terms: the sex ratio at birth, and the relative survival girls and boys. As infant mortality approaches zero, so does the second term, and the childhood sex ratio approaches the sex ratio at birth.

For a tractable empirical expression, we take Taylor series approximations ($ln(1 + x) \approx x$). Defining q_0 as overall infant mortality and $\mu = \frac{q_0^m - q_0^f}{q_0}$ as excess male mortality, we obtain:

$$(2) \quad ln(\frac{l_1^f}{l_1^m}) \approx ln(\frac{B^f}{B^m}) + \mu \cdot q_0$$

Equation (2) clarifies that infant mortality and excess male mortality combine to move the childhood sex ratio towards girls, away from the sex ratio at birth. The greater is excess male mortality, μ, the more that infant mortality skews the sex ratio among survivors. Importantly, this effect is roughly proportional to the level of infant mortality, so the effect will be negligible for populations with low infant mortality (e.g., rates below 20). However, the effect will be substantial in populations with high infant mortality. Absent extreme sex discrimination against girls, excess male mortality typically ranges from 15-30% (Hill and Upchurch 1995; Alkema et al. 2014). Starting from a healthy sex ratio at birth – say 5% more boys than girls – infant mortality rates of 150–300 could drive the sex ratio of the

one to four plays a negligible role in childhood sex ratios because it is much less male-skewed than infant mortality (Hill and Upchurch 1995, figure 1), meaning that ${}_5L_0$ broadly corresponds to an average of the l_1 values of each birth cohort.

surviving population to parity.

Of course, we would not expect a population with high infant mortality to have a healthy sex ratio at birth, as a growing body of work demonstrates that insults to maternal well-being push the sex ratio at birth towards females (e.g., Almond and Edlund 2007; Fukuda et al. 1998; Catalano 2003).[24] Male frailty, in utero and in early infancy, means that poor maternal-infant health will be reflected both in terms of fewer males being born, and fewer males surviving infancy. Thus the direct effect of infant mortality on childhood sex ratios will likely be reinforced by a female-tilted sex ratio at birth, as infant mortality and maternal health are closely linked (e.g., Kramer 1987).[25]

It follows that a structural interpretation (Goldberger 1973:5) of a regression of childhood sex ratios on infant mortality – equation (2) – would have maternal health as a latent variable. As explained by Goldberger, the estimated regression coefficients would correspond to "mixtures" of the underlying structural parameters. In our case, the estimated slope coefficient captures not only the direct effect of infant mortality on sex ratios (via excess male mortality), but also the correlation between the sex ratio at birth and the infant mortality, mediated through maternal health. For two populations with the same levels of infant mortality, we might then have different childhood sex ratios, depending on the degree to which the causes of this high infant mortality are related to maternal health.[26] In practice, we can expect different childhood sex ratios with a given level of infant mortality when populations differ in the terms of the nature and causes of that mortality. It follows that estimating structural parameters would be problematic. Fortunately, our focus is on making predictions, not estimating parameters. The extent to which childhood sex ratios

[24]The apparent mechanism is maternal stress hormones, which increase the probability of miscarriages, which are disproportionately male (James and Grech 2017, p. 51). The sex ratio at birth has been used as an indicator for maternal health and fetal loss (Davis, Gottlieb, and Stampnitzky 1998; Grech and Masukume 2016; Shifotoka and Fogarty 2013; Sanders and Stoecker 2015; Valente 2015; Guimbeau, Menon, and Musacchio 2021).

[25]Klasen (1994, p. 1064–1066) noted this relationship between sex ratio at birth and infant mortality in the context of 'missing women'.

[26]A key example is the extent of breastfeeding, which plays a vital role in infant mortality but can be unrelated to maternal health.

reflect, and therefore can predict, infant mortality is an empirical question, which we address with historical data from populations where both variables are available.

Data

To characterize the empirical relationship between childhood sex ratios and infant mortality, we assemble data from Europe, the US, and other settler societies, mostly from the mid-19th century onward.[27] Data for childhood sex ratios are taken from censuses or population registries, and for infant mortality are taken from official sources, International Historical Statistics (IHS), and the Human Mortality Database (HMD).

We pair a childhood sex ratio with an average rate infant mortality in preceding years.[28] We generally use the under-5 population, but other age-groupings yield the same basic results. The under-5 age group has a number of important advantages over younger ages.[29] First, it is more widely available from published sources. Second, the five-year age span increases the sizes of childhood populations, reducing the role of random variation in sex ratios.[30] Finally, pooling across ages reduces the impact of sex-biased age heaping. The starting points for our series are dictated by the availability of data. We end our series at the start of the 1960s; by then, rates of infant mortality in our sample populations were too low to materially affect childhood sex ratios, and ultrasound, which spread in the 1970s (Campbell 2013), was not yet a factor in sex-ratio patterns. We restrict our dataset to under-5 populations of at least

[27] See the data appendix for a fuller discussion of our sample. In brief, our non-US data cover: Sweden (1753–1960), Denmark (1836–1960), Belgium (1842–1960), England and Wales (1847–1961), the Netherlands (1855–1960), Scotland (1857–1960), New Zealand (1863–1961), Austria (1865–1961), Australia (1876–1961), Germany (1849–1961), Switzerland (1876–1960), Finland (1881–1960), Norway (1886–1960), France (1897–1954), Italy (1907–1961), and South Africa (1914–1921). For the US we have Massachusetts from 1856–1960, and then a growing number of states from 1900 onward.

[28] With some exceptions, we pair the under-5 sex ratio with the prior 5-year mean of the infant mortality rate. We have under-6 populations for 140 Prussian cases of 1890-1910. For Prussian districts in 1849 and some US states in 1900, we have just one year of infant mortality data (see the data appendix for details).

[29] At older ages, migration could become a significant factor.

[30] Random variation in sex ratios will not be small unless populations are large. To illustrate, model the sex proportion as binomial random variable, as in Visaria (1967, p. 33), with mean 1/2. With 10,000 children, the 90% CI is 6 percentage points, which is very large relative to the effects we seek to measure. With 50,000 children, the 90% CI shrinks to about 3 percentage points.

25,000. We have 571 observations for Europe and settler societies other than the US. For the US, we have 8 observations from the State of Massachusetts for the 19th century, and 177 observations from a variety of other aggregates (urban, rural, and mixed) for the years 1900 to 1940.[31] Thus a typical observation in our dataset pairs the (ln) under-5 sex ratio from a particular year with the average infant mortality rate for the preceding 5 years, for some country or sub-national unit (plotted below, in Figure 3). Summary statistics for these data can be found in the appendix.

Results

Childhood Sex Ratios Reveal Infant Mortality

We have established (above) the theoretical basis for childhood sex ratios reflecting infant mortality. In Figure 3 we plot under-5 sex ratios against infant mortality and their empirical correspondence is striking. High rates of infant mortality imply relatively more girls, and low rates relatively more boys. Moreover, the European and US data follow very similar patterns. Given this strong empirical relationship, childhood sex ratios can shed new light on infant mortality in populations lacking data on births or infant deaths.

We use the data in figure 3 to construct a range of plausible infant mortality rates given observed childhood sex ratios. We characterize the conditional distribution of sex ratios on infant mortality using quantile regression, allowing us to infer the likelihood of an observed sex ratio given hypothesized levels of infant mortality. Estimating a conditional quantile, we can then construct hypothesis tests, ruling out unlikely levels of infant mortality. In Figure 3 we plot the 10th percentile of the under-5 sex ratio conditional on infant mortality:

$$\hat{q}_{SR|IMR}(10\%) = \hat{\gamma} + \hat{\delta} \cdot IMR = -5.97 + 0.208 \cdot IMR$$

[31] By 1950, US infant mortality had fallen below 30 deaths per 1000, and state-level differences in white infant mortality were too small to be useful for our study.

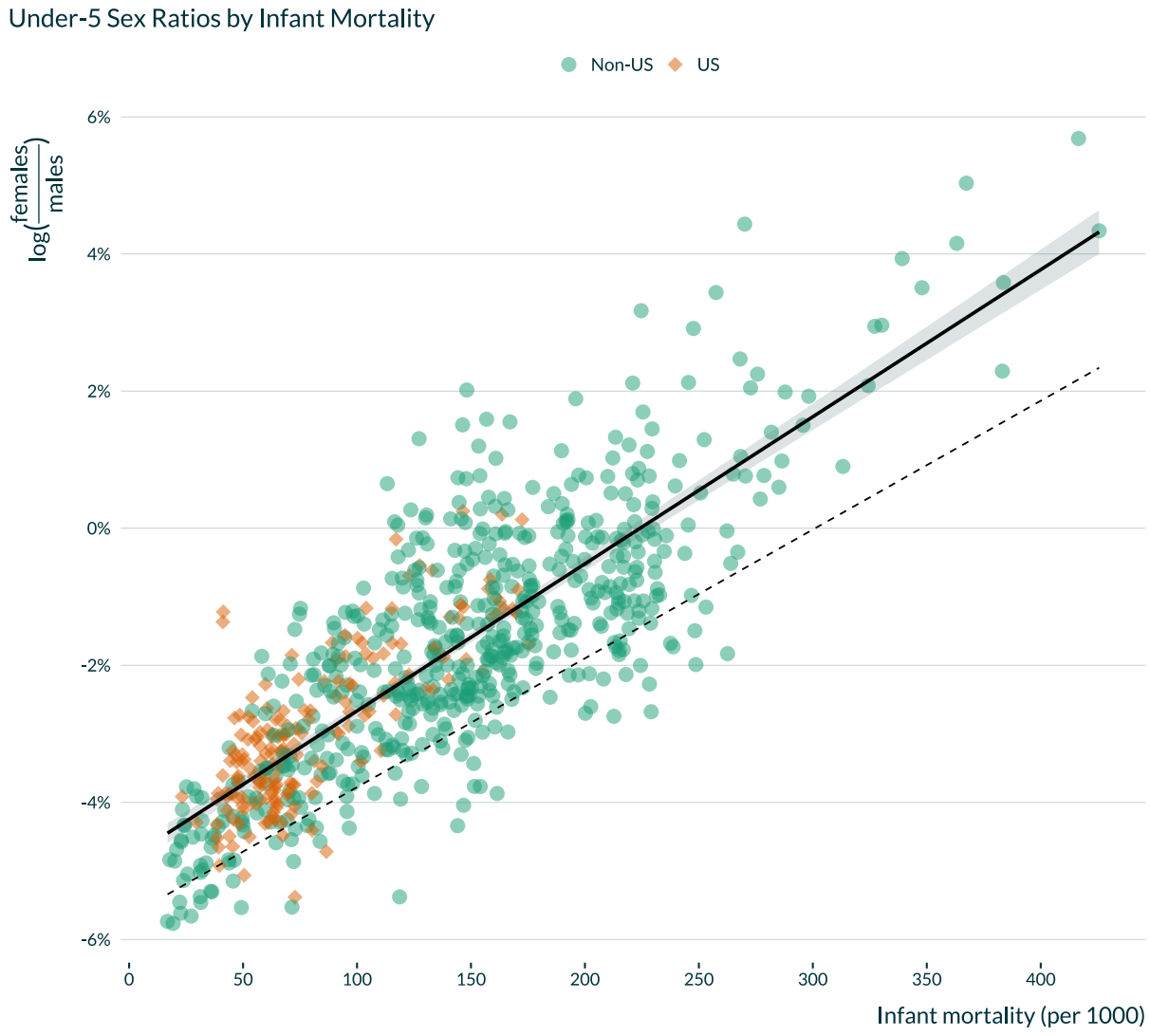

Figure 3: Infant mortality by under-5 sex ratios. The black line is the regression of under-5 sex ratios on infant mortality; the dashed line is the 10th percentile regression. See text below. Data mainly from Europe and the US (see Data section).

For an observed under-5 sex ratio of SR_i, we reject all infant mortality beyond the level which corresponds to this 10th percentile: i.e. reject if $IMR > \bar{IMR}$, where:

$$(3) \quad \bar{IMR} = \frac{SR + 5.97}{0.208}$$

Graphically, given an observed sex ratio, with 90% confidence we reject all infant mortality to the right of the dashed line plotted in Figure 3.

Beyond establishing probabilistic ranges, we can use equation (2) to infer infant mortality from childhood sex ratios. Equation (2) gives clear predictions for the relationship of childhood sex ratios and infant mortality. The relationship between the sex ratio and level of infant mortality should be roughly linear. In a regression of childhood sex ratios on infant mortality, the intercept coefficient should capture the healthy sex ratio at birth (5% male), and the slope coefficient the level of excess male mortality (15–30%).

We regress under-5 sex ratios (SR) on infant mortality (IMR) within our sample, based on the structural relationship established in equation (2).[32] The regression line is plotted in Figure 3:

$$\hat{SR} = \hat{\alpha} + \hat{\beta} \cdot IMR = -4.95 + 0.224 \cdot IMR$$

This estimated relationship closely conforms to the theoretical relationship above. The regression intercept (-4.95%) corresponds to a healthy sex ratio at birth, with about 5% more boys than girls, and the estimated slope coefficient (0.224) falls well within the usual range for 19th-century European populations (Hill and Upchurch 1995; Drevenstedt et al. 2008). Infant mortality explains more than two thirds of the variation of childhood sex ratios within our sample, with $R^2 = .68$. Inverting the expression above gives us a simple method

[32] We use simple bivariate regression, where regression weights are equal to the square root of the under-5 population. These are a-priori efficient, being proportional to one over the square root of the sampling variance of each observation. Our results are robust to other weights (see Figure 9).

for giving estimates of infant mortality from childhood sex ratios:[33]

$$(4) \quad \hat{IMR} = \frac{SR + 4.95}{0.224}$$

As proof of concept, we apply our prediction method (equation 4) to Massachusetts, which was the only US state with reasonably complete records on births and infant deaths going back to the mid-19th century. We drop the Massachusetts data, re-estimate our regression, and then predict IMR from under-5 sex ratios in Massachusetts. Plotted in Figure 4, we find a striking, if rough, correspondence between predicted infant mortality and the actual values (5-year averages). The Massachusetts example illustrates the promise of childhood sex ratio evidence for characterizing the approximate level of infant mortality in a population.

[33]These regression parameters are estimated with sampling uncertainty, so we calculate bootstrap prediction errors for $\hat{IMR}$. These errors are negligible relative to the levels of infant mortality we estimate (e.g., 95% CI of +/- 5 on predicted IMR of 100).

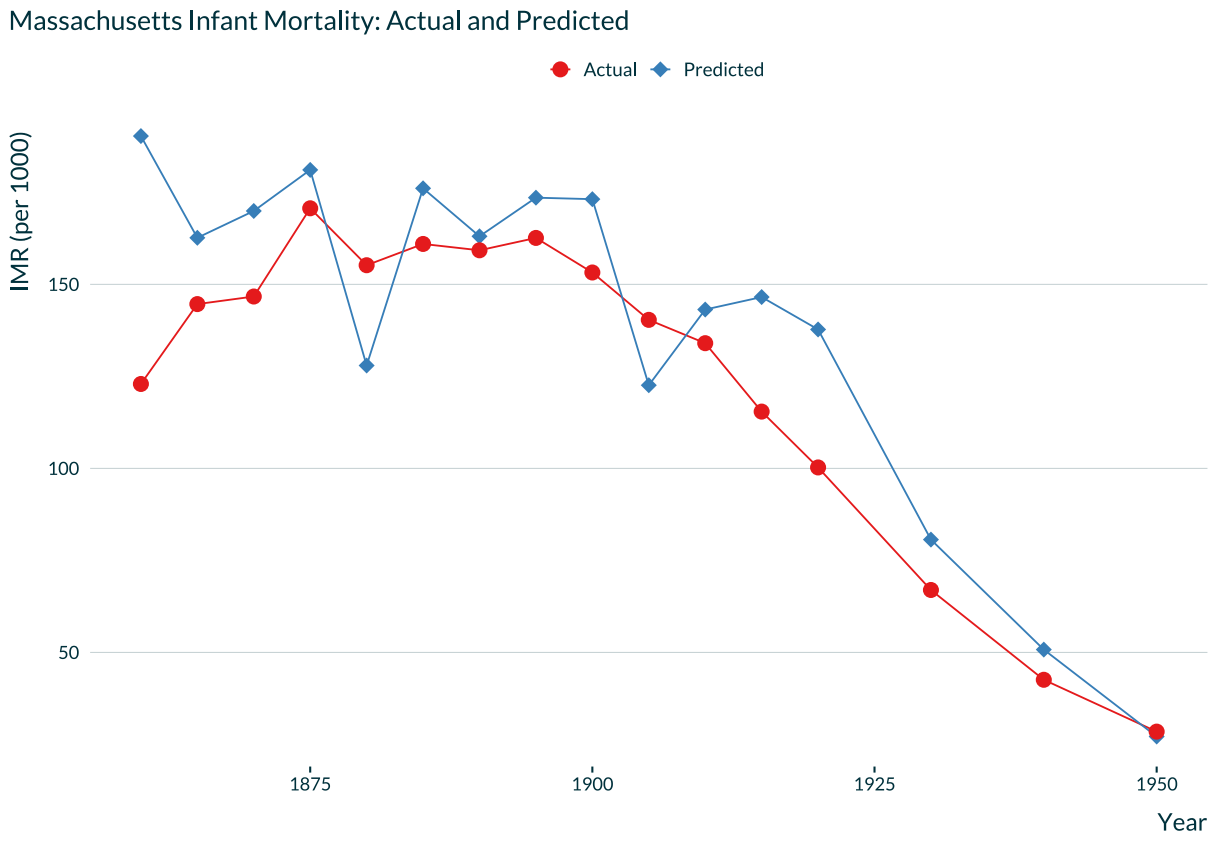

Figure 4: Out-of-sample prediction of infant mortality (5-year average) from childhood sex ratios: Massachusetts. Predicted values calculated from a regression of under-5 sex ratios on infant mortality. Regression data from Figure 4, excluding Massachusetts data. Massachusetts sex ratios from state and federal censuses; IMR from HSUS series Ab928.

US Infant Mortality 1850–1880

Having established the usefulness of childhood sex ratios for inferring infant mortality, we now turn to applying these methods to the 19th-century US white population. We draw on the four decennial censuses from 1850 to 1880, using data from both the published census volumes and the full count IPUMS samples.[34] We exclude the 1890 census because age-reporting in that year was inconsistent with practices in the rest of the censuses, biasing childhood sex ratios in 1890 toward males.[35]

As described above, the HSUS series places US white infant mortality in the period 1850 to 1880 at about 200 deaths per 1000. Such high infant mortality would have strong and simple implications for childhood sex ratios. Referring to our model above (equation 1), supposing a modest degree of excess male mortality – 20% – and a healthy sex ratio at birth – 5% more boys than girls – an infant mortality rate of 200 would result in a childhood sex ratio of parity. Referring to our simplest empirics – Figure 3 – we see that for populations with infant mortality of 200, under-5 sex ratios are similarly concentrated in the range of one percentage point on either side of parity.

The observed under-5 sex ratios of US whites baldly contradict these implications, ranging from 3.1% more boys than girls in 1870 to nearly 3.5% in 1850. Within our sample, such sex ratios are generally associated with infant mortality around 80 deaths per 1000, with 80% of observations falling between 50–120.

We plot our new estimates of IMR for US whites alongside those from Haines (1998) and

[34] We use the average of the IPUMS and published-census values, viewing both as plausible tallies of the underlying manuscripts. The two sources give very similar under-five sex ratios; 1850 shows the biggest discrepancy, with the full count IPUMS ratio 0.36% more male than the census volume's.

[35] The 1890 census recorded "age at nearest birthday" instead of "age at last birthday", which was used from 1850 to 1880, and from 1900 forward (US Census (1902: xlviii). A child approaching 5 years of age would be enumerated as age 5 in 1890, but in the other censuses they would be enumerated as age 4. Thus older 4-year-olds would be under-represented in the 1890 census under-five cohort (compared to the other censuses), biasing that cohort's sex ratio toward males (because the sex ratio among four-year-olds is less male than among infants, a result of excess male infant mortality). This pattern is evident in the census counts of the US-born populations of 1890 and 1900: the under-five cohort of 1890 numbered 6.49 million with a sex ratio 3.7% male; ten years later, the age 10 through 14 cohort numbered 6.65 million with a sex ratio 2.3% male. Further to this point, see the Robustness section below.

HSUS (2006) in Figure 5.[36] From the 1890s onward, our values line up well with existing estimates based on maternal recall (1895 and 1904) and vital statistics (1915 onward). But for 1880 and earlier – during which the HSUS series is based solely on extrapolation from age 5–20 mortality – our new estimates are much lower than those of HSUS.

For the period of 1850–1880, our point estimates of US infant mortality fall between 66–81 deaths per 1000. Building from our quantile regression above, which characterizes the 90th percentile of childhood sex ratios given infant mortality, we use equation (3) to construct an upper bound on US white infant mortality. At the 90% confidence level, we can consistently reject infant mortality approaching 140 deaths per 1000, with our upper bound ranging from 120 in 1850 to 137 in 1870. Thus we reject the HSUS values, which range from 167–218 across the period. We would also reject the hypothesis that US whites had infant mortality approaching that of, for example, England (IMR around 150) during the period.

[36] As discussed above, the HSUS series is sourced from Haines (1998) for the period up to 1910, and from official vital statistics for 1915 onward. In Figure 5, we correct the '1910' value from HSUS to 1904, and add the corresponding indirect estimate for 1900 (reflecting mortality circa 1894–1895) from Haines (1998, table A.IV).

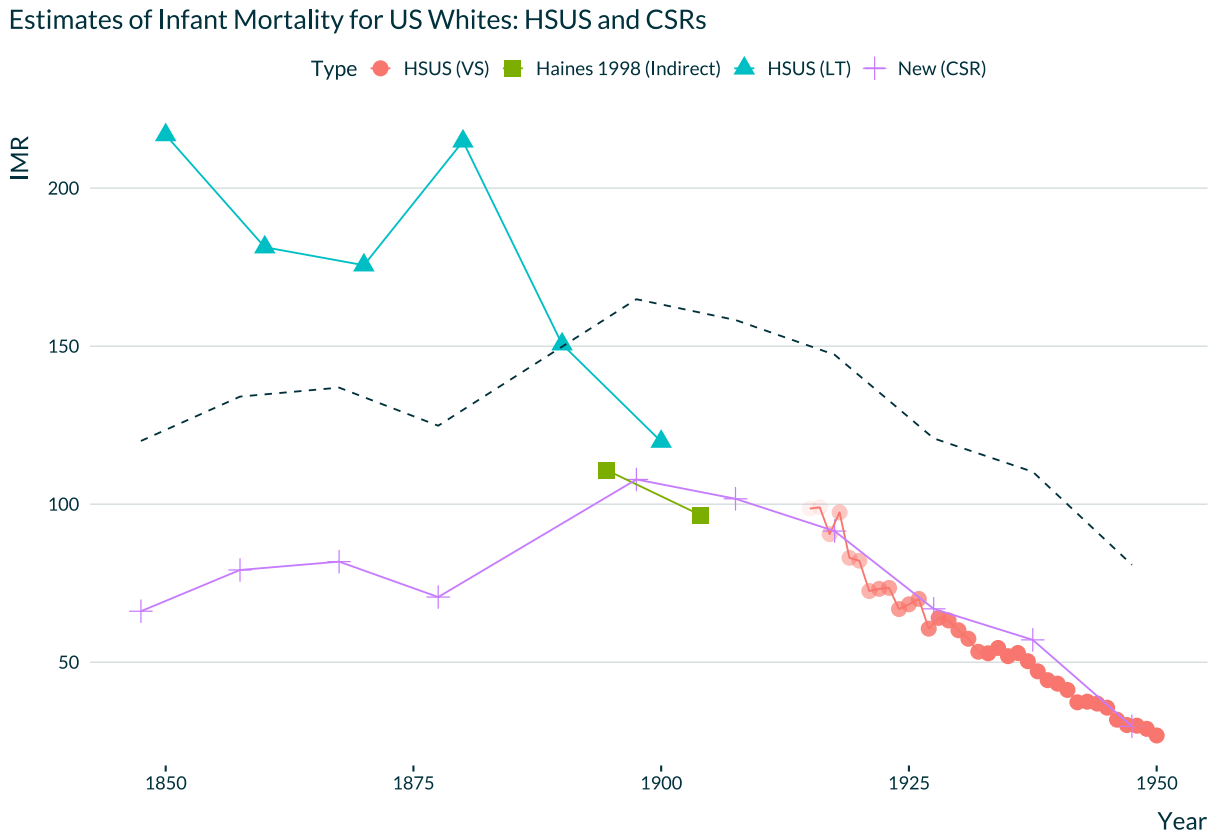

Figure 5: Estimates of US white infant mortality: 1850–1950. Plus-signs are our new estimates based on childhood sex ratios (using equation 4). The dashed line gives a 90% upper bound (using equation 3). As in Figure 1, the existing estimates from HSUS series Ab921 are broken up by type. The left segment (1850-1900) comes from life tables, the right segment from births and infant death records (shaded by degree of coverage). In the middle we plot the two indirect estimates from Haines (1998), for circa 1894–1895 (based on the 1900 census) and for circa 1904 (based on the 1910 census). The latter value appears in HSUS series Ab921 and Ab9 for the year 1910, but here we adjust it to the correct year.

Robustness

Among observed populations, under-5 sex ratios of around 3% more boys than girls are associated with relatively low rates of infant mortality (see above, Figure 3). In this sense, our qualitative result that US white infant mortality was relatively low in the 19th century is very robust, and is not sensitive to modifications to our empirical specification (such as allowing for non-linearity, alternative regression weights, or allowing the intercept to vary across countries; see Figure 9 in the appendix). Similarly, infant mortality rates around 200 deaths per thousand are associated with childhood sex ratios within 1% of parity, so the HSUS infant mortality values for the period 1850–1880 are simply inconsistent with the sex ratio evidence.

However, one key concern is the quality of the sex ratio data. Under-enumeration of young children is a common problem in historical censuses, including for the 19th-century US (e.g., Coale and Zelnik 1963, p. 10–11; Hacker 2013). If enumeration of young children was biased towards males, then observed childhood sex ratios would tend to understate the level of infant mortality.[37] Fortunately, we can test whether our results can be dismissed as artifacts of 19th-century enumeration problems.

We compare the under-5 sex ratio in one census to the age 10–14 sex ratio in the census ten years later – essentially following the cohort across the decade for a second measure of the under-5 sex ratio (a 'forward' measure). We look at the US-born white population of the nation as a whole so that immigration and inter-regional migration are not at play. The age 10–14 sex ratio promises to be a good proxy for the under-5 sex ratio ten years earlier: under-enumeration was much lower for ages 10–14 than the under-5 age group (Hacker 2013, figure 3), and child mortality after age four is generally both dramatically lower and less male-biased than infant mortality (Hill and Upchurch 1995). If a relative undercounting of infant girls in the 19th century is biasing our under-5 sex ratios toward boys – for a false

[37] We thank George Alter (personal communication) for both alerting us to this problem and suggesting how to address it.

impression of low infant mortality – then we should observe a relatively more female sex ratio among 10–14 year-olds ten years later.

Looking across the censuses, there is no evidence of relative underenumeration of young females in the 19th century. For each decennial census year from 1850 to 1940, Figure 6 plots the under-five sex ratio and the age 10–14 sex ratio from the next census (10 years later), for two measures of the under-five sex ratio (current and forward). The current and forward sex ratios line up very well, with similar trends and levels across both measures, with the sole exception of the census year 1890. There, the forward sex ratio measure is much less male than the current measure, powerfully signalling a male-biased enumeration of young children in 1890. However, as discussed above, that bias is an expected result of the anomalous age-question used in the 1890 census ('age at closest birthday' instead of 'age at last birthday').[38]

Thus the forward measure of childhood sex ratios strongly corroborates our basic results. Both measures show a population of some 3% more boys than girls in the mid-19th century, suggesting relatively low rates of infant mortality. Furthermore, Figure 6 shows a striking 'inverted-U' shape. Maternal-infant health appears to have deteriorated over the second half of the 19th-century, before beginning its well-documented improvement in the 20th. This fits well with the prevailing view that US population health deteriorated across much of the the 19th century.[39] However, it complicates any claims that the mortality transition emerged before the 20th century (e.g. Anderson, Charles, and Rees 2022, 128).

[38] As discussed above, see note 33. The age recording question meant that ages 4.5 and up were excluded from the under 5 cohort. For the 10–14 category, this age-recording issue is inconsequential, because the sex ratio at age 9 is very similar to at age 14. Among children, on the other hand, age 4 sex ratios are more female than age 0, because of excess male mortality.

[39] Fogel (1986), Pope (1992) and Hacker (2010) all find that life expectancy declined from 1800 to 1850, and Margo and Steckel (1983) and Komlos (1987) find that adult male heights declined over the same period.

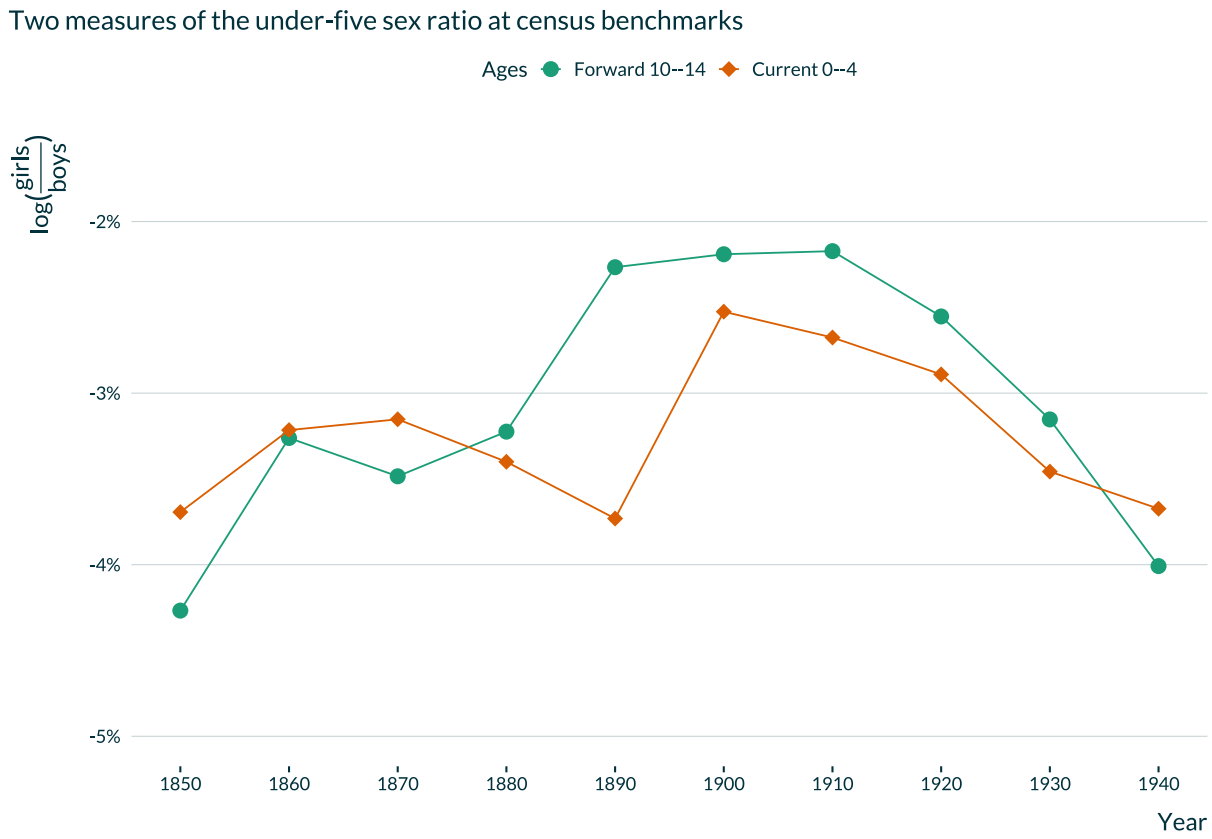

Figure 6: Two measures of the under-five sex ratio at census benchmarks, US native-born whites 1850–1940. The orange line connects the under-5 sex ratio in each census year. The green line connects the sex ratio of 10–14 year olds in the following census year (10 years later). As discussed in text, the 1890 under-5 sex ratio is male-biased due to the anomalous enumeration practices of the 1890 census. Data from IPUMS and census volumes.

Discussion

Childhood sex ratios allow us to overcome the challenge of the lack of vital statistics for the 19th-century US and characterize levels of infant mortality. With boys outnumbering girls by more than 3% at each of the decennial censuses from 1850 to 1880, we have clear evidence that US whites were a healthy population by the standards of the 19th century.[40] Figure 7 presents our estimates of the IMR for US whites alongside the HSUS (2006) series, against the backdrop of the well-documented IMRs of contemporary Europe. In sharp contrast to the HSUS series, we find US white infant mortality well below European levels before 1900. Our estimates for the 19th century are much different from those presented in HSUS, which come from life-table exercises. But our estimates line up well with the later values, which come from indirect and direct estimates of infant mortality.

Our estimates point to a dramatic revision of the evolution of infant mortality in the 19th-century US. According to the HSUS series, infant mortality plummeted in the closing decades of the 19th century, falling nearly in half.

We find the reverse, with infant mortality sharply higher circa 1900 and 1910 than it had been before 1880. Such a substantial rise in infant mortality during a period of tremendous economic growth – US per-capita income doubled between the 1875 to the 1910 (HSUS 2013, Series Ca11) – casts doubt on simple narratives of progress, such as the 'McKeown thesis.'[41] Instead, our results point to the challenges that modernization posed to population health. It was only after 1900 that the path of US infant mortality turned downwards, lining up with the advent of sanitation measures in US cities (Cain and Rotella 2008). This provides another piece of evidence for the growing consensus that investments in public health measures, rather than economic growth, drove the modern mortality transition.[42]

[40] As discussed above (Historical Background), this finding lines up well with a wide range of historical evidence on contemporary living standards.

[41] Although mostly disregarded in public health circles (Colgrove 2002), the 'McKeown thesis' (that economic growth drove mortality improvements in the 19th century; see McKeown 1976) continues to be evoked by economists (e.g., Anderson, Charles, and Rees 2022).

[42] See, e.g., Aykroyd and Kevany (1973); Preston and Van de Walle (1978); Cain and Rotella (2001);

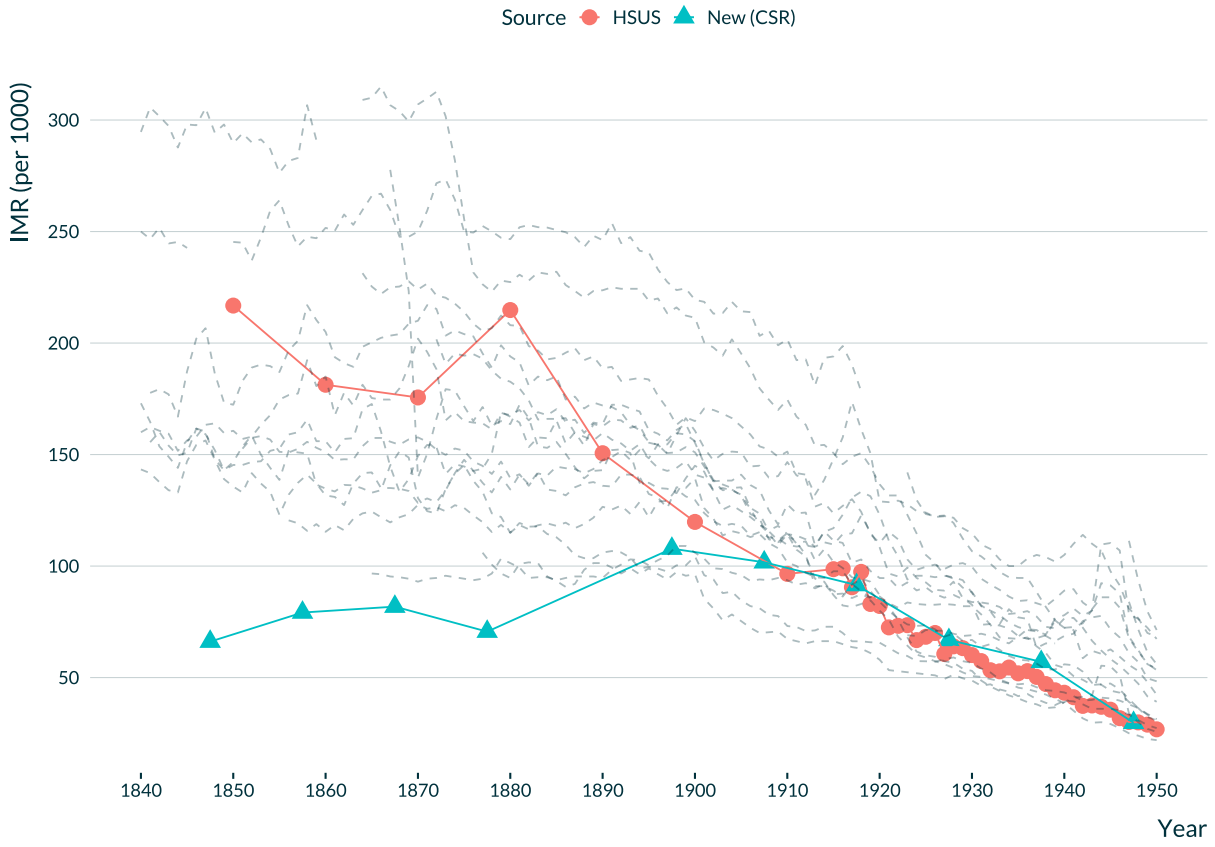

Figure 7: Estimates of US Infant mortality, plotted against a backdrop of European experiences (from Figure 1). Previous estimates comes from HSUS (2006, Ab921); new estimates are calculated as above, based on under-5 sex ratios, using equation 4.

This point is made abundantly clear when we disaggregate US whites into urban and rural populations. Using equation (4), we make estimates of infant mortality from childhood sex ratios, plotted in Figure 8. In line with existing work (e.g., Kearns 1988), we find a pronounced urban health penalty in the 19th century, with infant mortality some 80 to 100 points higher in urban than rural areas. After 1900 the urban penalty began to decline, and by 1930 it had almost disappeared, as found by Haines (2001, p. 47). It is clear from Figure 8 that the early-20th-century decline in infant mortality was primarily an urban phenomenon. As this was a period of rapid urbanization (the urban share of population increased from 35% in 1890 to 56% in 1930; U.S. Census Bureau 2012, table 10), the reduction in urban infant mortality was the critical factor for improvements in infant mortality overall.

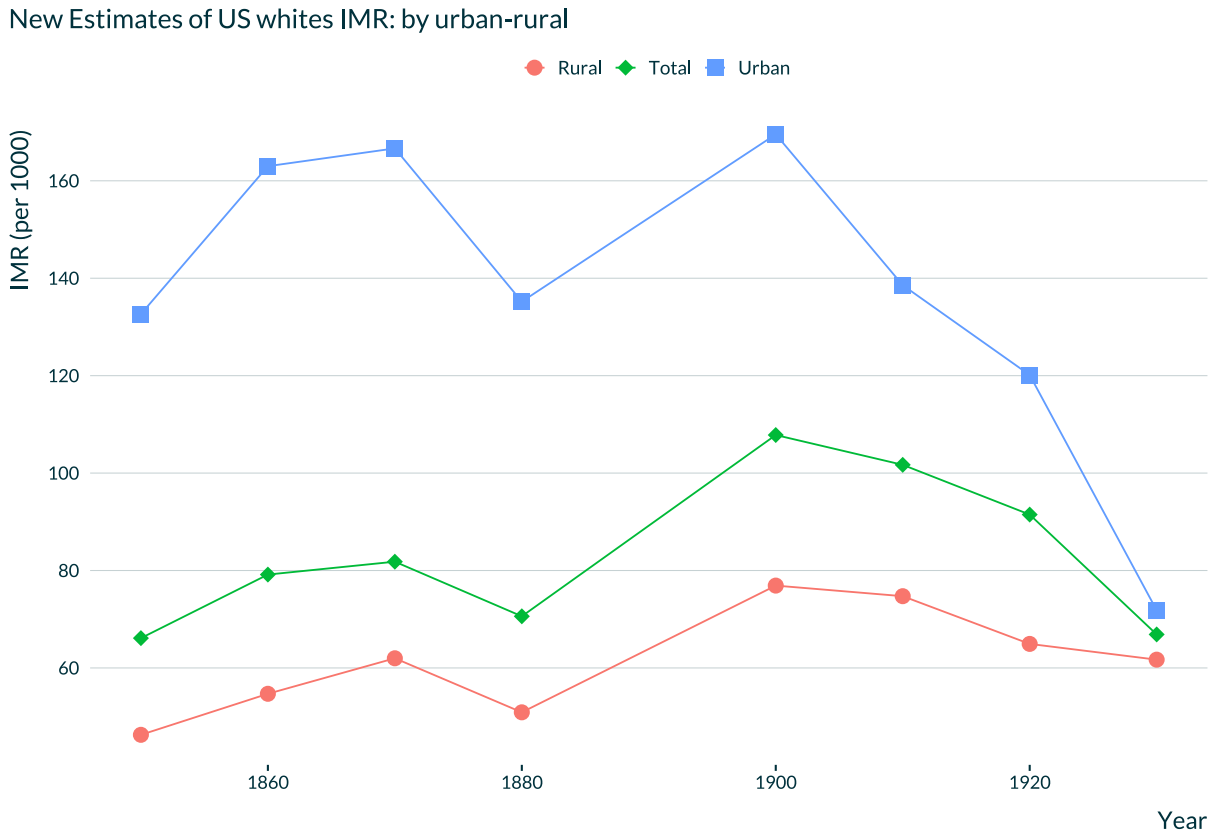

Figure 8: Under-5 sex ratio estimates infant mortality for US whites, urban vs. rural. Calculated via equation 4. Sex-ratio data from US census: published volumes and IPUMS.

Cutler and Miller (2005); Alsan and Goldin (2019).

Figure 8 also shows that the exceptional health of US whites in the mid-19th century was a rural phenomenon. While US urban infant mortality was comparable to contemporary London (around 150 deaths per 1000), rural US whites enjoyed levels of infant mortality well below any major European population. Under-5 sex ratios in the rural US place infant mortality under 60 deaths per 1000 circa 1850. For comparison, in the 19th century the healthiest Scottish counties had infant mortality above 80 (Lee 1991, table 1), and infant mortality in rural England was around 100 (Woods, Watterson, and Woodward 1988, p. 353). The combination of a mostly rural population (72% in 1880; U.S. Census Bureau 2012, table 10) with low infant mortality in rural areas, meant that US whites exceptionally health by 19th-century standards.

We suggest low inequality as a likely explanation for the good health of rural US whites. The US, especially the Midwest, was known for having relatively egalitarian access to land.[43] The importance of inequality is further supported by Ferrie's (2003, p. 36) finding that infant mortality in rural Illinois was half as much for families with some wealth as opposed to no wealth. We see nutrition as a key mechanism, with a more egalitarian distribution of land and income offering rural US whites superior diets to those of contemporary European peasants. More narrowly, maternal nutrition was likely the key channel by which low inequality was translated into good health, as it is a crucial determinant of infant mortality (e.g., Abu-Saad and Fraser 2010).[44]

Our proposed narrative is reminiscent of elements of the views of McKeown (1976) and Fogel (2004), insomuch as nutrition takes a leading role in determining historical mortality.

[43]See, for example, the discussion of Engerman and Sokoloff (2013). This access to land, of course, only applied to white settlers, as the land had been violently seized from indigenous communities (Carlos, Feir, and Redish 2022) – to say nothing of the enslaved Black population of the South (see discussion below).

[44]Breastfeeding was also likely very important. Salmon (1994) argues that breastfeeding past the first year was ubiquitous in the rural US, as opposed to parts of contemporary Europe, like Germany, where low breastfeeding contributed to high rates of infant mortality (Knodel and Van de Walle 1967). This raises a deeper question: to what extent was breastfeeding influenced by inequality? One possibility is that high inequality led to low breastfeeding, as women were forced to work longer hours and therefore unable to take the time needed to breastfeed their babies. This point has been made in the context of determinants of 21st-century US breastfeeding patterns (Wolf 2003).

However, these authors focused on the role of improved nutrition as caused by economic growth. We suggest, on the other hand, that adequate nutrition was already possible at pre-industrial levels of production, but was prevented by hierarchical social structures. We can only speculate to what extent better nutrition, and therefore health, might have prevailed in contemporary Europe under a more egalitarian distribution of land and income. Differences in infant mortality across 19th-century Europe were huge, sometimes larger than the total decline in infant mortality observed in most countries (e.g., Norway 100 vs. Germany 300). To what extent could these differences be explained by inequality?

Nowhere is the role of inequality more evident than in the US South, considering racial inequities under slavery. Childhood sex ratios provide clear evidence of relatively low infant mortality among 19th-century US whites, but they corroborate the most pessimistic views of Black infant mortality under slavery (McDevitt-Irwin and Irwin 2023). In 1850 and 1860, the under-five sex ratio of the slave population was remarkably skewed toward females, with over 2% more girls than boys, while among white children boys outnumbered girls by more than 3% (US Census 1860a). While at the extreme of our sample, the female-skewed childhood sex ratios of the enslaved suggest an infant mortality rate of 300 or more (McDevitt-Irwin and Irwin 2023, figure 1). Before the abolition of slavery, the 19th-century US featured an extreme contrast in terms of population health, with whites enjoying one of the lowest infant mortality rates in the world, while enslaved Blacks suffered one of the highest. The extremes of infant mortality found in the rural US at mid-century harshly illustrate the importance of social structures in determining population health, as well as the range of infant mortality possible in the pre-industrial era.

Conclusion

Infant mortality is a key indicator of historical population health and living conditions more generally. But until now, establishing even approximate levels of infant mortality for the

19th-century US has been an intractable problem due to a lack of data on births and infant deaths. Life table exercises (Haines 1979, 1998; Hacker 2010) have suggested a high rate of infant mortality for US whites: around 200 deaths per 1000 in the period 1850–1880. However, such values appear implausibly high in view of a range of other evidence and known patterns of historical infant mortality.

This paper provides a partial solution to the problem of a lack of data for standard estimates (direct or indirect) of infant mortality in the 19th-century US. We offer a new method for characterizing broad patterns of infant mortality, using childhood sex ratios from census data. Because of the well-known biological survival advantage of infant females, high rates of infant mortality tend to skew the surviving population towards girls. This theoretical relationship is strikingly evident in historical data from Europe and the US, providing a simple means to infer infant mortality rates from under-5 sex ratios. We use quantile regression to place bounds on plausible rates of infant mortality given observed sex ratios.

The US census reveals roughly 3% more males than females under the age of 5 for 19th-century US whites. These childhood sex ratios suggest that US white infant mortality in the period 1850–1880 was less than half of life-table based estimates: around 75 deaths per 1000, rather than 200. Using hypothesis testing, we reject at the 10% significance level an average infant mortality greater than 130 for US whites across period 1850–1880. Our results place US whites among the healthiest populations of the 19th century, with infant mortality substantially below levels found in Europe. The relative good health of US whites stood in sharp contrast to the experience of the Black population under slavery: childhood sex ratios suggest that Black infant mortality rates were some 250 points (4 times) higher than those of the white population.

On our evidence, the history of infant mortality in the US was not any simple variation on well-documented European patterns. In the 'pre-transition' period, US whites experienced much lower infant mortality than Europeans. Moreover, the 20th-century mortality decline

in the US was preceded by a substantial deterioration of maternal-infant health.[45] Rising infant mortality in the closing decades of the nineteenth century – a period of rapid economic growth and development – contradicts simple narratives of progress, like the 'McKeown thesis.'[46] Instead, our results point to the importance of public policy initiatives (e.g. water treatment) for overcoming the health challenges of mass urbanization.

Although refinements to our approach may offer more precise results in future applications, we have shown that in the absence of birth and death records, childhood sex ratios can provide plausible ranges of infant mortality solely from published census counts. Often available when vital statistics are not, childhood sex ratios can shed new light on maternal and infant well-being in historical populations, particularly during the 19th and 20th centuries.

[45] Even at its early 20th-century peak, US white infant mortality was still lower than in much of contemporary Europe (see Fig 7).

[46] This point echoes elements of Easterlin (1999), that economic growth alone did not lead to improved health, and Engerman (1997), that modern economic growth came with meaningful trade-offs to population well-being.

References

Abbott, Samuel W. 1897. "The Vital Statistics of Massachusetts: A Forty Years' Summary, 1856-1895." *Twenty-Eighth Annual Report of the Massachusetts State Board of Health.*

Abu-Saad, Kathleen, and Drora Fraser. 2010. "Maternal Nutrition and Birth Outcomes." *Epidemiologic Reviews* 32 (1): 5–25.

Alkema, Leontine, Fengqing Chao, Danzhen You, Jon Pedersen, and Cheryl C. Sawyer. 2014. "National, Regional, and Global Sex Ratios of Infant, Child, and Under-5 Mortality and Identification of Countries with Outlying Ratios: A Systematic Assessment." *The Lancet Global Health* 2 (9): e521–e530.

Almond, Douglas, and Lena Edlund. 2007. "Trivers-Willard at Birth and One Year: Evidence from US Natality Data 1983–2001." *Proceedings of the Royal Society B: Biological Sciences* 274 (1624): 2491–6.

Alsan, Marcella, and Claudia Goldin. 2019. "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880-1920." *Journal of Political Economy* 127 (2): 586–638.

Anderson, D. Mark, Kerwin Kofi Charles, and Daniel I. Rees. 2022. "Reexamining the Contribution of Public Health Efforts to the Decline in Urban Mortality." *American Economic Journal: Applied Economics* 14 (2): 126–57.

Aykroyd, W. R., and J. P. Kevany. 1973. "Mortality in Infancy and Early Childhood in Ireland, Scotland and England and Wales 1871 to 1970." *Ecology of Food and Nutrition* 2 (1): 11–19.

Beltrán Tapia, Francisco J., and Michail Raftakis. 2021. "Sex Ratios and Gender Discrimination in Modern Greece." *Population Studies*, 1–18.

Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz. 2017. "Sample-Selection Biases and the Industrialization Puzzle." *The Journal of Economic History* 77 (1): 171–207.

Bureau, U. S. Census. 2012. "Population and Housing Unit Counts." CPH-2-1. Washington, D.C.: U.S. Government Printing Office,

Cain, Louis, and Sok Chul Hong. 2009. "Survival in 19th Century Cities: The Larger the City, the Smaller Your Chances." *Explorations in Economic History* 46 (4): 450–63.

Cain, Louis, and Elyce Rotella. 2001. "Death and Spending: Urban Mortality and Municipal Expenditure on Sanitation." In *Annales de Démographie Historique*, 139–54. 1. Belin.

———. 2008. "Epidemics, Demonstration Effects, and Investment in Sanitation Capital by U.S. Cities in the Early Twentieth Century." In *Quantitative Economic History*. Routledge.

Caldwell, John C. 2001. "Population Health in Transition." *Bulletin of the World Health Organization* 79 (2): 159–60.

———. 2006. "Transmuting the Industrial Revolution into Mortality Decline." In *Demographic Transition Theory*, edited by John C. Caldwell, 157–80. Dordrecht: Springer Netherlands.

Campbell, S. 2013. "A Short History of Sonography in Obstetrics and Gynaecology." *Facts, Views & Vision in ObGyn* 5 (3): 213–29.

Carlos, Ann M., Donna L. Feir, and Angela Redish. 2022. "Indigenous Nations and the Development of the US Economy: Land, Resources, and Dispossession." *The Journal of Economic History* 82 (2): 516–55.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, eds. 2006. *Historical Statistics of the United States*. Millennium edition. Cambridge University Press.

Catalano, Ralph A. 2003. "Sex Ratios in the Two Germanies: A Test of the Economic Stress Hypothesis." *Human Reproduction* 18 (9): 1972–5.

Clarke, Joseph. 1786. "Observations on Some Causes of the Excess of the Mortality of Males Above That of Females." *The London Medical Journal* 9 (Pt 2): 179.

Coale, Ansley J., and Paul Demeny. 1983. *Regional Model Life Tables and Stable Populations*. 2nd ed. Studies in Population. Academic Press.

Coale, Ansley J., and Melvin Zelnik. 1963. *New Estimates of Fertility and Population in the United States*. Princeton University Press.

Colgrove, James. 2002. "The McKeown Thesis: A Historical Controversy and Its Enduring Influence." *American Journal of Public Health* 92 (5): 725–29.

Condran, Gretchen A., and Rose A. Cheney. 1982. "Mortality Trends in Philadelphia: Age-and Cause-Specific Death Rates 1870." *Demography* 19 (1): 97–123.

Condran, Gretchen A., and Eileen Crimmins. 1979. "A Description and Evaluation of Mortality Data in the Federal Census: 1850-1900." *Historical Methods* 12 (1): 1.

———. 1980. "Mortality Differentials Between Rural and Urban Areas of States in the Northeastern United States 1890." *Journal of Historical Geography* 6 (2): 179–202.

Costa, Dora L. 2015. "Health and the Economy in the United States from 1750 to the Present." *Journal of Economic Literature* 53 (3): 503–70.

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.

Davis, Devra Lee, Michelle B. Gottlieb, and Julie R. Stampnitzky. 1998. "Reduced Ratio of Male to Female Births in Several Industrial Countries: A Sentinel Health Indicator?" *Jama* 279 (13): 1018–23.

Davis, Kingsley. 1973. *Cities: Their Origin, Growth, and Human Impact*. San Francisco: W.H. Freeman and Company.

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

Easterlin, Richard A. 1999. "How Beneficent Is the Market? A Look at the Modern History of Mortality." *European Review of Economic History* 3 (3): 257–94.

Engerman, Stanley L. 1997. "The Standard of Living Debate in International Perspective: Measures and Indicators." In *Health and Welfare During Industrialization*, 17–46. University of Chicago Press.

Engerman, Stanley L., and Kenneth L. Sokoloff. 2013. "Five Hundred Years of European Colonization: Inequality and Paths of Development." In *Settler Economies in World History*, 65–103. Brill.

Ferrie, Joseph P. 2003. "The Rich and the Dead. Socioeconomic Status and Mortality in the United States, 1850-1860." In *Health and Labor Force Participation over the Life Cycle: Evidence from the Past*, 11–50. University of Chicago Press.

Fogel, Robert W. 1986. "Nutrition and the Decline in Mortality Since 1700: Some Preliminary Findings." In *Long-Term Factors in American Economic Growth*, 439–556. University of Chicago Press.

———. 2004. *The Escape from Hunger and Premature Death, 1700-2100: Europe, America, and the Third World*. Vol. 38. Cambridge University Press.

Fogel, Robert W., Stanley L. Engerman, Roderick Floud, Gerald Friedman, Robert A. Margo, Kenneth Sokoloff, Richard H. Steckel, T. James Trussell, Georgia Villaflor, and Kenneth W. Wachter. 1983. "Secular Changes in American and British Stature and Nutrition." *Journal of Interdisciplinary History* 14 (2): 445.

Fukuda, Misao, Kyomi Fukuda, Takashi Shimizu, and Henrik Møller. 1998. "Decline in Sex Ratio at Birth After Kobe Earthquake." *Human Reproduction (Oxford, England)* 13 (8): 2321–2.

Goldberger, Arthur S. 1973. "Structural Equation Models: An Overview." *Structural Equation Models in the Social Sciences*, 1–18.

Grech, Victor, and Gwinyai Masukume. 2016. "The Sex Ratio at Birth in South Africa May Be a Sentinel Health Indicator." *Early Human Development* 103: 225–27.

Grech, Victor, Charles Savona-Ventura, and P. Vassallo-Agius. 2002. "Unexplained Differences in Sex Ratios at Birth in Europe and North America." *BMJ: British Medical Journal* 324 (7344): 1010–1.

Guimbeau, Amanda, Nidhiya Menon, and Aldo Musacchio. 2021. "Short and Medium-Run Health and Literacy Impacts of the 1918 Spanish Flu Pandemic in Brazil." No. w26929. National Bureau of Economic Research.

Hacker, J. David. 2010. "Decennial Life Tables for the White Population of the United States, 1790–1900." *Historical Methods* 43 (2): 45–79.

———. 2013. "New Estimates of Census Coverage in the United States, 1850-1930." *Social Science History; Durham* 37 (1): 71–101.

Hacker, J. David, Martin Dribe, and Jonas Helgertz. 2023. "Wealth and Child Mortality in the Nineteenth-Century United States: Evidence from Three Panels of American Couples, 1850." *Social Science History* 47 (3): 333–66.

Haines, Michael R. 1977. "Mortality in Nineteenth Century America: Estimates from New York and Pennsylvania Census Data, 1865 and 1900." *Demography* 14 (3): 311–31.

———. 1979. "The Use of Model Life Tables to Estimate Mortality for the United States in the Late Nineteenth Century." *Demography* 16 (2): 289–312.

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

———. 2001. "The Urban Mortality Transition in the United States, 1800-1940." In *Annales de Démographie Historique*, 33–64. 1. Belin.

———. 2006. "Vital Statistics." In *Historical Statistics of the United States*, 4th ed., A: Population:381–90. New York: Cambridge Univ Press.

Haines, Michael R., and Hallie J. Kintner. 2000. "The Mortality Transition in Germany, 1860; Evidence by Region." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 33 (2): 83–104.

Haines, Michael R., and Samuel H. Preston. 1997. "The Use of the Census to Estimate Childhood Mortality: Comparisons from the 1900 and 1910 United States Census Public Use Samples." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 30 (2): 77–96.

Hill, Kenneth. 1995. "The Decline of Childhood Mortality." In *The State of Humanity*, edited by Julian L. Simon, 37–50. Oxford: Blackwell.

Hill, Kenneth, and Dawn M. Upchurch. 1995. "Gender Differences in Child Health: Evidence from the Demographic and Health Surveys." *Population and Development Review*, 127–51.

James, William H., and Victor Grech. 2017. "A Review of the Established and Suspected Causes of Variations in Human Sex Ratio at Birth." *Early Human Development* 109: 50–56.

Kearns, Gerry. 1988. "The Urban Penalty and the Population History of England." In *Society, Health and Population During the Demographic Transition*, 213–36. Umeå Universitet.; Demografiska Databasen No. 4. Stockholm: Almqvist & Wiksell.

Kesztenbaum, Lionel, and Jean-Laurent Rosenthal. 2017. "Sewers' Diffusion and the Decline of Mortality: The Case of Paris, 1880." *Journal of Urban Economics* 98: 174–86.

Klasen, Stephan. 1994. "'Missing Women' Reconsidered." *World Development* 22 (7): 1061–71.

Knodel, John, and Etienne Van de Walle. 1967. "Breast Feeding, Fertility and Infant Mortality: An Analysis of Some Early German Data." *Population Studies* 21 (2): 109–31.

Komlos, John. 1987. "The Height and Weight of West Point Cadets: Dietary Change in Antebellum America." *The Journal of Economic History* 47 (4): 897–927.

Komlos, John, and Marieluise Baur. 2004. "From the Tallest to (One of) the Fattest: The Enigmatic Fate of the American Population in the 20th Century." *Economics & Human Biology* 2 (1): 57–74.

Kramer, M. S. 1987. "Determinants of Low Birth Weight: Methodological Assessment and Meta-Analysis." *Bulletin of the World Health Organization* 65 (5): 663–737.

Kunze, Kent. 1979. "The Effects of Age Composition and Changes in Vital Rates on Nineteenth Century Population: Estimates from New Data." PhD thesis, The University of Utah.

Lee, Clive H. 1991. "Regional Inequalities in Infant Mortality in Britain, 1861-1971: Patterns and Hypotheses." *Population Studies*, 55–65.

Lynch, Katherine A., Geraldine P. Mineau, and Douglas L. Anderton. 1985. "Estimates of Infant Mortality on the Western Frontier: The Use of Genealogical Data." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 18 (4): 155–64.

Maconochie, Noreen, and Eve Roman. 1997. "Sex Ratios: Are There Natural Variations Within the Human Population?" *BJOG: An International Journal of Obstetrics & Gynaecology* 104 (9): 1050–3.

Margo, Robert A., and Richard H. Steckel. 1983. "Heights of Native-Born Whites During the Antebellum Period." *The Journal of Economic History* 43 (1): 167–74.

McDevitt-Irwin, Jesse, and James R. Irwin. 2023. "Racial Health Inequities in the 19th Century US: Evidence from Childhood Sex Ratios."

McKeown, Thomas. 1976. *The Modern Rise of Population*. New York: Academic Press.

Mitchell, Brian. 1998. *International Historical Statistics: Europe 1750-1993*. 4th ed. New York: Grove's Dictionaries, Inc.

Pope, Clayne L. 1992. "Adult Mortality in America Before 1900: A View from Family Histories." In *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*, 267–96. University of Chicago Press.

Preston, Samuel H., and Michael R. Haines. 1991. *Fatal Years: Child Mortality in Late Nineteenth-Century America*. Princeton University Press.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishing.

Preston, Samuel H., and Etienne Van de Walle. 1978. "Urban French Mortality in the Nineteenth Century." *Population Studies* 32 (2): 275–97.

Riley, James C. 2005a. "The Timing and Pace of Health Transitions Around the World." *Population and Development Review* 31 (4): 741–64.

———. 2005b. "Estimates of Regional and Global Life Expectancy, 1800-2001." *Population and Development Review* 31 (3): 537–43.

Salmon, Marylynn. 1994. "The Cultural Significance of Breastfeeding and Infant Care in Early Modern England and America." *Journal of Social History*, 247–69.

Sanders, Nicholas J., and Charles Stoecker. 2015. "Where Have All the Young Men Gone? Using Sex Ratios to Measure Fetal Death Rates." *Journal of Health Economics* 41 (May): 30–45.

Sawyer, Cheryl Chriss. 2012. "Child Mortality Estimation: Estimating Sex Differences in Childhood Mortality Since the 1970s." *PLOS Medicine* 9 (8): e1001287.

Sen, Amartya. 1992. "Missing Women." *BMJ: British Medical Journal* 304 (6827): 587.

Shifotoka, Anna, and Andrew W. Fogarty. 2013. "Is the Human Sex Ratio at Birth a Sentinel Health Indicator That Is Associated with HIV/AIDS and Tuberculosis Prevalences in the 21st Century?" *J Epidemiol Community Health* 67 (4): 327–31.

Struyck, Nicolaas. 1740. *Inleiding Tot de Algemeene Geographie, Benevens Eenige Sterrekundige En Andere Verhandelingen*. Amsterdam: Issak Tirion.

Théré, Christine, and Jean-Marc Rohrbasser. 2006. "Facing Death in the Early Days of Life: Inequality Between the Sexes in Enlightenment Demographic Thought." *The History of the Family* 11 (4): 199–210.

Thompson, Amanda L. 2021. "Greater Male Vulnerability to Stunting? Evaluating Sex Differences in Growth, Pathways and Biocultural Mechanisms." *Annals of Human Biology* 48 (6): 466–73.

United Nations Department of Economic and Social Affairs, Population Division. 2022. "World Population United Nations Department of Economic and Social Affairs, Population Division Prospects 2022: Summary of Results." UN.

U. S. Census Office, US Census. 1902. *Twelfth Census of the United States, Taken in the Year 1900 ... Volume II. Population Part II.* Vol. 2. Census Reports ... Twelfth Census of the United States, Taken in the Year 1900. Washington, D.C.: Government Printing Office.

Valente, Christine. 2015. "Civil Conflict, Gender-Specific Fetal Loss, and Selection: A New Test of the Trivers-Willard Hypothesis." *Journal of Health Economics* 39: 31–50.

Visaria, Pravin M. 1967. "Sex Ratio at Birth in Territories with a Relatively Complete Registration." *Eugenics Quarterly* 14 (2): 132–42.

Waldron, Ingrid. 1998. "Sex Differences in Infant and Early Childhood Mortality: Major Causes of Death and Possible Biological Causes." In *Too Young to Die: Genes or Gender?*, edited by Alan D. Lopez and Lado T. Ruzicka, 64–83. 155. New York: United Nations.

Wargentin, Pehr. 1755. "Anmärkningar Om Nyttan Af årliga Förtekningar På Födda Och Döda I et Land." *Sjette Stycket. Kongl. Vetenskaps Academiens Handlingar För år 1755* 16: 241–53.

Weir, David. 1997. "Economic Welfare and Physical Well-Being in France, 1750-1990." In *Health and Welfare During Industrialization*, 161–200. University of Chicago Press.

Williamson, Jeffrey G. 1982. "Was the Industrial Revolution Worth It? Disamenities and Death in 19th Century British Towns." *Explorations in Economic History* 19 (3): 221–45.

Wolf, Jacqueline H. 2003. "Low Breastfeeding Rates and Public Health in the United States." *American Journal of Public Health* 93 (12): 2000–2010.

Woods, Robert. 1993. "On the Historical Relationship Between Infant and Adult Mortality." *Population Studies* 47 (2): 195–219.

Woods, Robert I., Patricia A. Watterson, and John H. Woodward. 1988. "The Causes of Rapid Infant Mortality Decline in England and Wales, 1861–1921 Part I." *Population Studies* 42 (3): 343–66.

Appendix

Summary Statistics

Table 1: Summary Statistics: Data for Regressions

Year	pre-1849	1849–1869	1870–1899	1900–1929	1930–1961
N	41	52	263	283	117

	Min.	1st. Qu.	Median	Mean	3rd Qu.	Max.
Sex Ratio (% F/M)	-5.77	-3.30	-2.06	-1.96	-0.86	5.68
IMR (per 1000)	17	72	131	133	174	426
Population	25,324	89,951	162,941	385,239	324,037	17,358,552

Robustness

Here we present the IMR predictions of various regressions of infant mortality on childhood sex ratios. The 'base' specification, used throughout the paper, uses all data from Figure 3, is least-squares, and weighted by the square root of the under-5 population. We conduct several robustness checks. First we allow the intercept to differ for each country. Then we allow for different weights (unweighted, and weighted by total population). Finally, we allow for a non-linear relationship between sex ratios and infant mortality (a cubic spline with three knots). All of the results are broadly similar, and agree with our basic qualitative result: 19th-century US infant mortality was much lower than previously thought.

US Under-5 Sex Ratios

We use under-5 sex ratios for US whites throughout this paper. The data are an average of IPUMS and published census volume values.

Table 3: US Under-5 Sex Ratios

Year	Rural	Total	Urban
1850	-3.915	-1.981	-3.471
1860	-3.726	-1.301	-3.178
1870	-3.563	-1.219	-3.119
1880	-3.812	-1.923	-3.370
1900	-3.229	-1.154	-2.537
1910	-3.278	-1.848	-2.674
1920	-3.497	-2.261	-2.902
1930	-3.570	-3.343	-3.454

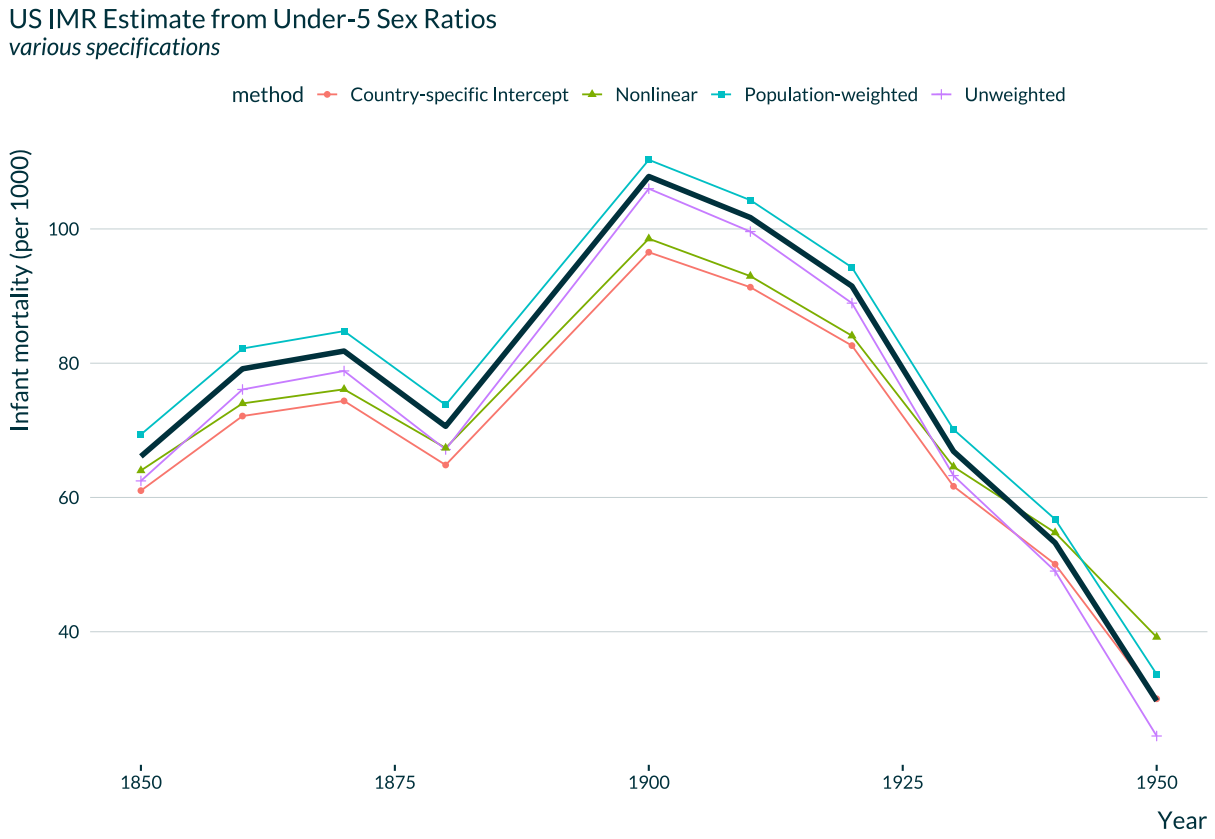

Figure 9: Alternative estimates of US white IMR, based on sex ratios. The black line is our main specification, used throughout in the text, which is least-squares weighted by the square root of population.

Software Used

Analysis done in R version 4.3.1 (2023-06-16), with the following packages:

Table 4: R Packages

Package	Loaded version	Date	Source
dplyr	1.1.2	2023-04-20	CRAN (R 4.3.0)
forcats	1.0.0	2023-01-29	CRAN (R 4.3.0)
ggplot2	3.4.2	2023-04-03	CRAN (R 4.3.0)
kableExtra	1.3.4	2021-02-20	CRAN (R 4.3.0)
lubridate	1.9.2	2023-02-10	CRAN (R 4.3.0)
mediocrethemes	0.1.3	2023-05-10	Github (vincentbagilet/mediocrethemes)
purrr	1.0.1	2023-01-10	CRAN (R 4.3.0)
quantreg	5.95	2023-04-08	CRAN (R 4.3.0)
readr	2.1.4	2023-02-10	CRAN (R 4.3.0)
SparseM	1.81	2021-02-18	CRAN (R 4.3.0)
stringr	1.5.0	2022-12-02	CRAN (R 4.3.0)
tibble	3.2.1	2023-03-20	CRAN (R 4.3.0)
tidyr	1.3.0	2023-01-24	CRAN (R 4.3.0)
tidyverse	2.0.0	2023-02-22	CRAN (R 4.3.0)

Data Sources

Categories of "race" follow usage in the source, unless otherwise noted.

Sources for Figures & Analyses

Sources for Figure 1 (Infant Mortality Rates, 1840-1990)

US white IMR are from HSUS (2006) Series Ab921. In this series, IMR values at decennial census benchmarks 1850-1910 are from Haines (1998: 163-65, 167). As discussed in our text, the values for 1850-1900 come from Haines' model life tables (1979:307; 1998:158-59), based on census data on age 5-20 mortality and population by age. The value for 1910 (Haines 1998:167) is an indirect estimate of the IMR, based on census data on population-by-age and children ever-born and surviving (maternal recall). Although presented for 1910 in the HSUS series, Haines reports the value for circa 1904 (as discussed in our text). For 1915 to 1990, the Series Ab921 data are annual, based on vital statistics (registrations of births and of infant deaths). These annual data are for the "Birth Registration Area" (BRA) which covered about 1/3 of the US population in 1915 and expanded over time, reaching coverage of the entire US in 1933 (HSUS Series Ab33). Figure 1 background IMR are for European populations, from IHS (2013), except for England & Wales and Scotland which are from the UK Office of National Statistics. Austria (1840-1990), Belgium (1840-1990), Denmark (1840-1990), Finland (1866-1990), France (1840-1990), Germany (1840-1937), West Germany (1946-1989), East Germany (1946-1989), Ireland (1864-1990), Italy (1863-1990), Netherlands (1840-1990), Norway (1876-1990), Switzerland (1871-1990), Sweden (1840-1990): IHS Series A7.
England & Wales (1850-1990), Scotland (1855-1990): UK ONS Vital Statistics Annual (downloaded 2023.0927)

Sources for Figure 2: Infant mortality by age 5–20 mortality.

Rates of infant and age 5-20 mortality are HMD estimates, from HMD life tables (Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org; data downloaded on 2022 July 17).
The data cover Australia (1921, 1925), Belgium (1841 and quinquennially 1845-1910, 1920, 1925), Canada (1921, 1925), Denmark (quinquennially 1835-1925), England and Wales (1841 and quinquennially 1845-1925), Finland (1878 and quinquennially 1880-1925), France (1816 and quinquennially 1820-1925), Italy (1872 and and quinquennially 1875-1925), Netherlands (quinquennially 1850-1925), Norway (quinquennially 1850-1925), New Zealand (1901 and quinquennially 1905-1925), Scotland (quinquennially 1855-1925), Spain (1908 and quinquennially 1910-1925), Sweden (1751 and quinquennially 1755-1925), and Switzerland (1876 and quinquennially 1880-1925).
The age 5-20 mortality rates for the shaded band in Figure 2 are from the life tables of Haines (1998:156-65); those life tables are the source of the 19th-century infant mortality rates in HSUS series Ab921.

Data sources for Figure 3 (Under-five sex ratios by infant mortality) and for regression analysis of the CSR:IMR relationship

The dataset for Figure 3, and used for regression analysis, is comprised of highly credible data on infant mortality rates and on childhood sex ratios. These data are direct estimates of infant mortality, taken from vital statistics, combined with under-5 sex ratios from censuses or population registries. We have national-level data for Sweden (1757–1960), Denmark (1840–1960), Belgium (1846–1960), the Netherlands (1859–1960), Scotland (1861–1960), New Zealand (1867–1961), Australia (1880–1961), Switzerland (1880–1960), Finland (1885–1960), Norway (1890–1960), France (1901–1954), Italy (1911–1961), South Africa (1918–1921), Germany (1920–1960), England & Wales (1926-1961), and Austria (1930-1961).
Our sub-national data includes Prussian districts (1849–1910), divisions of England & Wales (1851-1921), districts of Bavaria (1863–1880), Austrian Provinces (1865–1910), the State of Massachusetts (1860–1915), and various aggregates within the US (1900–1930). More specific information, including sources, follows.

For many historical populations, the *Human Mortality Database* provides access to official statistics for infant mortality rates and under-five sex ratios. We expand our geographic scope by also drawing on vital statistics and census data from various official sources for populations not included in the HMD.[47] In many cases, the national-level data are available from *International Historical Statistics* (Palgrave Macmillan (Ed.) 2013), which we abbreviate as *IHS* below. Specific sources and methods by polity follow.

Australia (1876–1961) Infant mortality rates for 1876–1901 are from McDonald et al. (1987:58).[48] Rates for 1901–1961 are from Australian Bureau of Statistics, Historical Population.[49] Under-5 populations by sex are census values for non-aboriginal populations. We have decennial data from 1881–1921 and single-year values for 1933, 1947, 1954, and 1961. The data for 1881 and 1891 are reported in Caldwell (1987:33–34). The 1901 and 1911 data are from the 1911 Census of Australia.[50] Data for 1921, 1933, 1947, 1954, and 1961 are reported in the Census of 1966.[51]

Austria Austrian Provinces (1865–1910) Infant mortality rates for 1865-1880 are calculated from births and infant deaths, reported annually in issues of Austria's *Statistisches Jahrbuch.*[52] Data for 1886-1910 are reported annually in the volumes of *Österreichische*

[47] The HMD is restricted to national populations "where death registration and census data are virtually complete" (HMD Overview). We include cases where the data are less "complete", requiring data only for infant mortality rates and childhood sex ratios. We also data from sub-national aggregates (the HMD has national data).

[48] Series MFM 154, in Chapter 3 of Vamplew (1987), *Australians – Historical Statistics.*

[49] Deaths data downloads, Table 5.4 "Infant mortality rates, states and territories, 1901 onwards", released 2019-04-18. Downloaded 2021-06-21

[50] *Census of the Commonwealth of Australia taken for the night between the 2nd and 3rd April, 1911*, Vol. II, Part 1 – Ages, pp. 10-11.

[51] Commonwealth Bureau of Census and Statistics (1970), *Census of Population and Housing, 30 June 1966 Commonwealth of Australia.* Volume 1. Population: single characteristics, part 1. Age, pp. 10–11.

[52] E.g. the 1865 data are in *Statistisches Jahrbuch der Österreichischen Monarchie - Für das Jahr 1866* (Wien, 1868), pp. 18, 20-21. The Jahrbuch issues, whose titles vary somewhat, are available from austrian literature online.

Statistik, Bewegung der Bevölkerung ....[53] For Provinces of Austria, we have under-5 populations by sex for 1869, 1880, 1890, 1900, and 1910, from Statistics Austria.[54] **Austria, national data** (1930-1961) Infant mortality rates (1930–1961) are from *IHS* (2013: 3577,3580,3583), Series A7. Under-5 populations by sex are for the years 1934, 1951, and 1961, reported in Statistik Austria, *Statistisches Jahrbuch 2010*.[55]

Belgium (1842–1961)

Infant mortality rates for Belgium (1842–1961) are HMD estimates (downloaded 2021-10-26). Under-5 populations by sex are census data, decennially 1846–1866 and 1880–1910, with single-years 1930, 1947, and 1961. The data were obtained through the HMD (downloaded 2021-07-01). The data for 1846, 1856, 1866, 1880, 1890, 1900, and 1910 are reported in the volumes for 1893, 1908, and 1923–24 of *Annuaire Statistique de la Belgique*[56]. The source for the 1930 data is the 1940 volume of *Annuaire Statistique de la Belgique et du Congo Belge* (pp. 34–35). The 1947 data are from the 1947 census of Belgium.[57] The data for 1961 are from the 1961 census.[58]

Denmark (1836–1960) Infant mortality rates (1836–1960) are HMD estimates (downloaded on 2021-10-26). Under-5 populations by sex are quinquennial 1840–1860 and 1900–1960, and decennial 1870–1890. The data were obtained through the HMD (downloaded on 2021-07-01). The Danish censuses until 1901 were taken as of February 1 of the census year (Andreev 2002:14-15), and so we take the childhood sex ratios from those census data as measures for the prior year. Similarly, the population data for 1906 onward refer to populations of January 1, and we use those for measures of the prior year's sex ratio. The data for 1840-1900[59] are from Danmarks Statistik (1905), *Befolkningsforholdene i DK i det 19. Aarhundrede*, STATISTISK TABELVÆRK, FEMTE RÆKKE, LITRA A NR. 5, Tabel 46, p. 55; available online According to the HMD (DNKref.pdf), the data for 1901–1960 are "population estimates ... produced by Danmarks Statistik", which were "obtained directly from the statistical office."

England & Wales

English sub-national data (1846-1921)

[53]For example, the 1886 data are in *Österreichische Statistik, Bewegung der Bevölkerung der im Reichsrathe vertretenen Königreiche und Länder im Jahre 1886*. The volumes for 1886–1890, 1896–1900, and 1906–1910 are available online in the Österreichische Statistik, 1880- section of the Österreichische Nationalbibliothek.

[54]STATcube – Statistical Database of STATISTICS AUSTRIA, Dataset: Population census data since 1869 by age and Provinces, downloaded 2023-02-20.

[55]2.08 Bevölkerung 1869 bis 2001 nach fünfjährigen Altersgruppen und Geschlecht (Population 1869 to 2001 by five-year age groups and sex, p. 45

[56]For 1846, 1893:64; for 1856, 1909:64; and 1926:30 for 1866 and decennially 1880–1900. These volumes are available online from HathiTrust: 1893, 1908, and 1923–24)

[57]Institut National de Statistique (1951), *Recensement Général de la Population, de L'Industrie et du Commerce au 31 décembre 1947, tome V, Répartition de la population par âge*, Tableau 1 - Répartition des habitants par âge et sexe ..." (p. 10). Bruxelles: Imprimerie Fr. Van Muysewinkel. The volume is available online from KU Leuven libraries.

[58]Institut National de Statistique (1965). *Recensement Général de la Population, 31 décembre 1961, tome V, Répartition de la population par âge.* Bruxelles: Institut National de Statistique. 1965); available online from KU Leuven libraries.

[59]census years 1841-1901

We have under-five populations for the eleven Registration Divisions of England decennially from 1851 to 1911. For 1921 the data are for individual or grouped Administrative Counties (see below).

Infant mortality rates for the eleven Registration Divisions are calculated from births and infant deaths in the Annual Reports of the Registrar-General until 1910 (specific page references are available in our replication datafiles). The 1911 child-sex ratio data are paired with an average infant mortality rate for the years 1906-1910 because 1911 data were not available.[60]

From 1851 to 1911 the census dates in England were on or near April 1, so about 1/4 of the census year had elapsed. Accordingly, our 5-year average infant mortality rates were constructed to reflect 1/4 of the census year, 3/4 of the year 5 years prior, and the full years in between.

Infant mortality rates for 1916-1921 are for Administrative Counties and County Boroughs, reflecting the change in 1911 from Registration to Administration areas (Seventy-Fourth Annual Report of the Registrar General (1911), pp. vii-viii) for vital statistics reporting. Infant mortality rates are calculated from births and infant deaths in the Registrar General Annual Reports, 1916-1921. With the census referring to the population as of June 20, 1921, for the prior five-year infant mortality rate we give 0.45 weight to 1921, 0.55 to 1916, and 1 to each of 1917-1920.

Under-five populations by sex for the eleven Registration Divisions are from the following publications: Census of Great Britain, 1851, Population Tables, I, Numbers of the Inhabitants, Report and Summary Tables (London 1852), pp cxcii; Census of England and Wales for the year 1861, Population Tables, Vol. II, "Ages, Civil Condition, . . . " (London 1863), p, xiv (Summary Tables, Table II); Census of England and Wales, 1871, Population Abstracts, "Ages, Civil Condition, . . . " (London 1873), p. xvi (Summary Tables, Table II); Census of England and Wales, 1881, Volume III, "Ages, Condition as to marriage, occupations . . . " (London 1883), pp. 3, 31, 81, 125, 165, 215, 277, 319, 375, 425, 463; Table 1 for each of the eleven Registration Divisions, in the "Divisional Tables" of Volume 3 of the 1891 Census of England and Wales ("Ages, Condition as to Marriage, . . . " (London: 1893), pp. 3, 29, 85, 137, 177, 223, 289, 329, 399, 453, 491); Census of 1901, Summary Tables, Table XXVIII "Ages of persons, males and females, in registration divisions and counties" (1903: pp. 162-171); Census of England, 1911, Vol. VII, Ages and Condition as to Marriage, Table 11 (London: 1913, pp. 312-373).

The 1921 data (for Administrative Counties and County Boroughs, see above). Under-5 populations by sex are from the 1921 Census (General Tables, Table 37 (pp. 145-150)).

[60] The 1911 data on births and infant deaths were not published for the Registration Districts; starting in 1911, vital statistics reporting shifted from Registration areas to Administrative areas (1911 Annual Report of the Registrar-General, pp. vii-viii). The 1911 census data was April 2; to approximate the average infant mortality from April 2, 1906 to April 2, 1911 we take a weighted average of 1906-1910, weighting 1906 by 0.8 and the other four years by 1.05.

Smaller counties or boroughs are aggregated with adjacent units for under-5 populations over 30 thousand, referring to the "Geographical Divisions" in the 1921 census (General Tables, Table 33, pp. 140-41 (refer to R code and replication datafiles for our aggregates). For England in 1921 we have 33 observations, including 6 urban areas (Birmingham, Leeds, Liverpool, London, Manchester, and Sheffield).

English national data (1922–1961) Infant mortality rates (1922–1961) are from the ONS Dataset Vital statistics in the UK: births, deaths and marriages(downloaded 2021-09-27). –> Under-5 populations by sex for England and Wales are quinquennial for 1926–1961, from the Historic Mortality Datasets of the National Archives.[61] Five-year average values of the IMR are paired with the under-5 sex ratios. **Finland** (1881–1960) Infant mortality rates (1881–1960) are HMD estimates (downloaded on 2021-10-26). Under-5 populations by sex are quinquennial from 1885 to 1960, obtained through the HMD (downloaded 2022-02-28). The HMD identifies Statistics Finland as the source of the data.[62] **France** (1897–1954) Infant mortality rates (1897–1968) are HMD estimates (downloaded on 2021-10-26). Under-5 populations by sex are quinquennial 1901–1946, with single-years 1954, 1962, 1968. The data were obtained through the HMD (downloaded on 2021-07-01), which identifies the source as Vallin & Meslé (2001).[63]

German Republic (1920–1933)
Infant Mortality Rates (1921–1933) are from IHS (2013: 3577, 3580), Series A7. Under-5 populations by sex are census values for 1925 and 1933; the data are from the *Statistisches Jahrbuch* of 1929 and 1939.[64] IHS (2013:3454, Series A2) also reports these age-sex population data, but rounded to the nearest thousand.[65]

West Germany (1956-1960)
Infant mortality rates (1956–1960) are HMD estimates (downloaded on 2021-10-26). Under-5 populations by sex for 1960 were obtained through the HMD (downloaded on 2021-10-26), which identifies the source as Statistisches Bundesamt.[66]

[61] RG 69/2, Historic Mortality: 1901–1995 dataset, Population, 1901–1995 (file POPLNS.csv), downloaded 2021-06-18.

[62] Under-five populations for 1885–1940 and 1945–1970 were received as computer files by the HMD from Statistics Finland: "Population estimates for years 1866–1940," and "Population estimates for years 1941–1995." This according to the "Data Sources" (https://mortality.org/hmd/FIN/DOCS/ref.pdf – login required) on the Finland page of the HMD website (accessed 2022-03-02.)

[63] The "Data sources" (https://mortality.org/hmd/FRATNP/DOCS/ref.pdf – login required) on the HMD data page for France describe the source as follows: "Vallin, J. and F. Meslé. (2001). Tableau I-C-1: Population par sexe et âge (de 0 à 100 ans), au 1 janvier, de 1899 à 1998, avec deux estimations selon le territoire pour les années de changement de territoire [revised post-publication]. In: Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle. Paris: Institnational d'études démographiques.cite Table Tableau I-C-1: Population par sexe et âge (de 0 à 100 ans), au 1 janvier, de 1899 à 1998" (accessed 2022-03-03).

[64] The 1925 data from *1929*, p. 14; 1933 from *1939*, p. 14.

[65] The IHS value for 1933 differs from ours; we use the value from the 1933 census (June 16); the IHS values for 1933 are consistent with the estimates for Dec. 31, 1933, found in *Statistisches Jahrbuch 1936*, p. 12.

[66] Annual population estimates as of December 31st, by age (0–94, 95+) and sex. Unpublished data.

Italy (1907–1961) Infant mortality rates (1907–1961) are from Istat Time Series.[67] Under-5 populations by sex for 1936 and decenially 1911–1931 and 1951–1961, from Istat (Italian National Institute of Statistics), Time Series.[68]

Kingdom of Bavaria (1863-1880)

Infant mortality and under-5 population by Regierungsbezirk. We have infant mortality rates for 1863-80 and under-5 sex ratios for 1867, 1875, and 1880. Infant mortality data for 1862-1875 are from Mayr (1878), *Die Bewegung der bayerischen Bevölkerung in den Jahren 1862/63 bis 1875.*[69] Infant mortality data for 1876-80 are from *Zeitschrift des Königlich-Bayerischen Statistischen Bureaus.* 13. 1881.[70] The 1867 census of Bavaria has under-5 populations by sex.[71] The under-5 data for 1875 are from *Die bayerische Bevölkerung nach Geschlecht, Alter, Civilstand und Staatsangehörigkeit: Volkszählung von 1875.* The 1880 census data for under-5 populations by sex are from Beiträge zur Statistic Bayerns, vol. 45-46 (1882-1883).

Kingdom of Prussia (1849, 1871-1910)

We have data at the level of the Regierungsbezirk (district).[72] We have under-five populations by sex, for the years 1849, 1875, and 1880. We have under-six populations by sex quinquennially from 1895 to 1910. All but the 1849 data are from the "Galloway Prussia Database 1861 to 1914". That database also provides infant mortality data (births and infant deaths) annually for 1871-1910.

With the exception of 1849, we pair under-5 sex ratios with the 5-year rolling means of infant mortality. We do not have Prussian infant mortality data for 1845-48, and so we pair the single year of infant mortality data for 1849 with the under-five sex ratio for that year. The 1849 data are from *Tabellen und amtliche Nachrichten über den Preußischen Staat für das Jahr 1849*; Vol. 1 for under-five populations, Vol. 2 for births and infant deaths.

New Zealand (1863–1961)

Infant mortality rates are for the non-Maori population from 1863–1945 and for the total population from 1947–1960. Data for 1863–1936 are from Stats NZ Store House.[73] The data for 1936–1945 are from The New Zealand Official Year-book 1957.[74] Data for 1947–1961 are for the total population (including Maori), from Stats NZ Inforshare.[75] Under-5 census populations by sex are for 1867, 1874, and 1881; quinquennially for 1886–1926 and 1951–1961; and also for 1936 and 1945. Data are for the non-Maori population until 1951. The data for 1867, 1874, and 1881 are found in the 1881 census.[76] Quinquennial data for

[67] Health, Infant mortality rate by age at death and sex; perinatal mortality rate by sex - Years 1863-2013 (Table_4.8.xls).

[68] Population, Population by age class and sex, aging ratio and dependency ratio at Census from 1861 to 2011 according to reference year borders (Table_2.2.1.xls).

[69] https://www.digitale-sammlungen.de/view/bsb11362367?page=12,13

[70] p. 191 for births, p. 198 for infant deaths.

[71] *Die Volkszählung im Königreiche Bayern vom 3. December 1867. 2: Die bayerische Bevölkerung nach Alter, Civilstand und Geschlecht.*

[72] We exclude the very small Sigmaringen from our data set; all the other Regierungsbezirke have u5 populations over 25 thousand.

[73] A2.7 Infant mortality rate and infant mortality number (spreadsheet), Thorns/Sedgwick non-Maori (column 3).

[74] Section 4 – Vital Statistics. European Infant Mortality.

[75] Population, Death Rates - DMM, Infant mortality rate (Annual-Dec).

[76] *Results of a census of the colony of New Zealand, taken for the night of the 3rd of April, 1881*, Chapter

1886–1916 are reported in the 1916 census.[77] The data for 1936, 1945, and quinquennially 1951–1961, are from the Stats NZ Store House.[78]

Netherlands (1855–1960)
Infant mortality rates (1855–1960) are HMD estimates (downloaded on 2021-10-26). Under-5 populations by sex (1859, 1869, and quinquennially 1875–1970) were obtained through the HMD (downloaded on 2021-07-01), which identifies the sources as the NIDI mortality database for 1859–1949 and Statistics Netherlands (Centraal Bureau voor de Statistiek) for 1950–1960.
Norway (1886–1960)
Following Backer (1961), we deem credible IMR data for Norway to start with the year 1876.[79] Infant mortality rates (1886–1970) are from IHS (2013: 3578, 3581, 3585); Statistics Norway online data on births and infant-deaths corroborate the IHS infant mortality data.[80] Under-five populations by sex are census values, decennially 1890–1930 and 1950–60, and for the year 1946.[81] Data for 1890–1900 are from Statistics Norway (1910).[82] Data for 1910–1930 are reported in the 1930 census.[83] The rest of the age-sex data for Norway are taken from published census volumes from the respective years: 1946 from Statistics Norway (1951), *Folketellingen 1946, Hefte 3*[84]; 1950 from Statistics Norway (1953), *Folketellingen 1950, Hefte 2.*[85]; 1960 from Statistics Norway (1963), *Folketellingen 1960, Hefte 2.*[86]; and 1970 from Statistics Norway (1971)[87](https://www.ssb.no/a/histstat/nos/nos_a448.pdf) (Population by age and marital status 31 December 1970), pp. 24–25.].

28, Table 1, "Showing the Increase of Persons of Both Sexes, Males, and Females (exclusive of Maoris), at different Ages, in the Intervals between the various Censuses, from December, 1864, to April, 1881."

[77] *Results of a census of the Dominion of New Zealand . . . 1916*, Part II Ages, p. 1.

[78] Spreadsheet (182.xls) titled A1.6 Population by age and sex (Long-term data series; Population;), spreadsheet A1.6 (citing Bloomfied (1984), "Census Reports: Table II.6. Age Groups . . . 1874-1976").

[79] Although counts of births and infant deaths start with the year 1836, we are guided by the judgment of Julie E. Backer, writing as "former chief of the Population Statistics Division, Central Bureau of Statistics of Norway". According to Backer (1961, p. 36), until 1876 infants who died early inflated counts of the stillborn, with live-births and infant deaths correspondingly understated. STATISTISK SENTRALBYRÅ (Oslo 1961). Although some early publications from Statistics Norway report IMR data from before 1876, their *Historical Statistics* of 1978, 1994, and 2000 present 5-year average values of IMR starting with 1876. In our view, that corroborates our conclusion that 1876 marks the start of reliable IMR data for Norway.

[80] Statistisk sentralbyrå, Historisk statistikk, 3.13 Folkemengde, fødte, døde, ekteskap, flyttinger og folketilvekst.

[81] The census values refer to January 1 of a year so we treat them as the prior year's ending value (so our 1890 U5 counts are from the January 1, 1891 census). The IHS and HMD list Norway's population data with the census years (so our 1890 value is listed in HMD as 1891).

[82] *Norges Folkemængde fordelt paa de enkelte aldersaar, 1846-1901*, Norges Officielle Statistik. V. 113, pp. 32, 34.

[83] Statistics Norway (1934), Folketellingen 1930, Hefte 5. *Folkemengden fordelt efter kjønn, alder og ekteskapelig stilling*, p. 2.

[84] *Folkemengden etter kjønn, alder og ekteskapelig stilling, . . .*, Tabeller p. 2.

[85] *Folkemengden etter kjønn, alder og ekteskapelig stilling . . .* (Population census December 1, 1950, Second volume, Population by sex, age, and marital status . . .), Tabeller p. 2.

[86] *Folkemengden etter kjønn, alder og ekteskapelig status.*

[87] *Folkemengden etter alder og ekteskapelig status 31. desember 1970*

Scotland (1857–1961)
Infant mortality rates (1857–1971) are HMD estimates (downloaded on 2021-10-26). Under-5 populations by sex are decennial 1861–1901 and quinquennial from 1911 to 1971; the data were obtained through the HMD (downloaded on 2021-07-01); original sources are as follows. The quinquennial data for 1861 to 1881 are published in the 1881 census.[88] Data for 1891–1901 are in the 1901 census.[89] Quinquennial data for 1911 to 1936 are from the General Register Office for Scotland.[90] Quinquennial data for 1941 to 1971 are from General Register Office for Scotland.[91]
South Africa (1913–1921)
Infant mortality rates (1913–1921) are from *IHS* (2013:219) Series A7. We have under-5 census populations by sex for 1918 and 1921, reported in the 1922 and 1925 volumes of the *Official Yearbook* of South Africa.[92]

Sweden (1753–1960)
Infant mortality rates (1753–1960) are from Statistics Sweden.[93] We have under-5 populations by sex for 1757, 1763, 1850, and quinquennially for 1785-1805, 1815-1835 and 1860-1970. Data for 1860-1970 are from Statistics Sweden.[94] For years before 1860, we use "official" counts reported by Sundbärg (1908:180).[95] We use years for which those "official" counts are consistent with Sundbärg's "corrected" counts (pp. 208, 216, 224), in terms of childhood sex ratios; the latter figures are used by the HMD.[96]
Switzerland (1875–1960)
Infant mortality rates (1875–1960) are calculated from data on births and infant-deaths from Historical Statistics of Switzerland, Marriage, Birth, and Death.[97] These IMRs are

[88]Scotland Census Office (1883), *Ninth decennial census of the population of Scotland . . . 1881 . . . Vol. II*, Appendix tables; with the 1861 and 1871 data in Table XXII, "Population of Scotland in 1861 and 1871, in sexes and ages . . . " (p. xxxii) and the 1871 and 1881 in Table XXI, "Population of Scotland in 1871 and 1881, in sexes and ages . . . " (p. xxxii). The volume is available online from HathiTrust.

[89]Scotland Census Office (1903), *Eleventh decennial census of the population of Scotland . . . 1901 . . . Vol II*, Appendix Tables, Table 1, "Population of Scotland in 1891 and 1901, distinguishing males and females at each year of life . . . " (p. xxxii). Available online from Google Books.

[90]Mid-year population estimates by sex and five year age group, 1911-1938. The HMD reports these as "Retrieved 15 May 2008" http://www.gro-scotland.gov.uk.

[91]Mid-year population estimates by sex and single year of age until the last age 85+ (1939-1970) or 90+ (1971-2001); unpublished data received by HMD via email on 28 February 28, 2007.

[92]The 1918 data are in Union office of census and statistics (1923), *Official Yearbook of the Union and of Basutoland, Bechuanaland Protectorate and Swaziland, No. 5 –1922* (pp. 158–59); Pretoria: The Government Printing and Stationary Office. The 1921 data are in Union office of census and statistics (1927), *Official Yearbook of the Union and of Basutoland, Bechuanaland Protectorate and Swaziland, No. 8 –1925* (p. 868); Pretoria: The Government Printing and Stationary Office.

[93]Statistical Database, Population, Population statistics, Deaths, Live births, stillbirths and infant mortality rates by sex. Year 1749–2020 (accessed 2023-09-15).

[94]Statistical Database, Population, Population statistics, Number of inhabitants, Population by age and sex. Year 1860–2021 (accessed 2022-02-28). The HMD uses these data.

[95]We relied on a variety of internet translation sites to access Sundbärg's tables and discussion, which are in Swedish.

[96]We deem two counts to be consistent when their child sex ratios differ by less than 0.5% (log basis). When the difference is greater, we deem the observations to be unreliable.

[97]HSSO, 2012. Tab.C.41. hsso.ch/2012/c/41 (Total Deaths (Excluding Stillborn Births) by Age Group

corroborated by *IHS* (2013: 3578,3582) Series A7. We have under-5 populations by sex for 1880, 1888, decennially 1900–1930, 1941, and decenially 1950-1960. The data are from Historical Statistics of Switzerland, Population.[98]

United States of America

Except as otherwise noted, we use the 20th-century US vital statistics definition of urban, referring to cities with population 10,000 or more.

The State of Massachusetts (1860–1925)

We use state totals quinquennially 1860-1895, 1905-1915, and 1925. We do not use the state's totals for 1900, 1920, and 1930; for those years we use various regional breakdowns of Massachusetts data (see below). The state-level data are for the total population (white and nonwhite).[99] Infant mortality rates (1856–1925) for the state are from *HSUS (2006)* Series Ab928.[100] Massachusetts state censuses provide under-five populations by sex decennially 1865-1925.[101] The US federal censuses include the state's data decennially for 1860-1890 and 1910.[102] We average the values from published federal census volumes with the available IPUMS full count data (1860-1880, 1910)[103]

Other states and areas of the US (1900, 1920, 1930, 1940)

US areas in 1900 include 23 observations. These are comprised of rural Northern New England (ME, NH, VT); rural Southern New England rural (CT, MA, RI); Boston MA, other MA urban, other New England urban; NY rural, Brooklyn NY, Manhattan NY, other New York City, other NY urban; NJ rural, NJ urban; Philadelphia PA, other PA cities (registration cities with population over 4,000); MI rural, MI urban; Cleveland & Cincinnati;

1867–1995) and HSSO, 2012. Tab.C.5a hsso.ch/2012/c/5a (Marriage, Birth, and Death 1867–1995: General Overview).

[98]HSSO, 2012. Tab. B.8a. hsso.ch/2012/b/8a (Total Residential Population by Age in Five Year Increments (Approximate Ages), 1860–1990)

[99]The nonwhite population of Massachusetts was too small to affect the patterns of interest and appropriate vital statistics (births and infant deaths) often are not available by race.

[100]The data for 1856-1941 are from Massachusetts vital statistics; after 1942, data are from US vital statistics.

[101]*Abstract of the Census of Massachusetts, 1865*, p. 2; *The census of Massachusetts: 1875, Volume I, Population and social statistics*, p. 269 (the published total for age-one females corrected from 15589 to 13589 via pp. 263-68); *The census of Massachusetts: 1885, Volume I, Population and social statistics, Part 1*, p. 434; *Census of the Commonwealth of Massachusetts: 1895, Volume II, Population and social statistics*, p. 422; *Census of the Commonwealth of Massachusetts 1905, volume 1, population and social statistics*, p. 480; *The decennial census 1915*, p. 478. These are available online

[102]*Ninth Census, Volume II, The Vital Statistics of the United States*, Table XXIII, pp. 563, 575 (data for 1860 as well as 1870). *Statistics of the population of the United States at the tenth census (June 1, 1880)*, Table XXI, p. 592. *Report on the population of the United States at the eleventh census: 1890, Part II*, Table 3, pp. 104–105. *Twelfth census of the United States, taken in the year 1900, Population Part II* (Census Reports Volume II), Ages, Table 3, pp. 110–111. *Thirteenth census of the United States taken in the year 1910, volume 1, population 1910, General Report and Analysis*, Table 43, p. 380.

[103]Steven Ruggles, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. Minneapolis, MN: IPUMS, 2021. The 1890 census manuscripts have not survived, so there is no full count data for that year.

Chicago; Milwaukee & Minneapolis & St Paul; St Louis; other Midwestern cities (registration cities with population over 4,000); registration cities of the South; registration cities of the West. For 1900, aggregates were formed to achieve a minimum under-5 population over 49,000 in order to reduce the role of random variation in sex-ratio data.

Infant mortality rates are single-year values calculated from births and infant deaths reported in US Census Office (1902), Twelfth Census, Census Reports Volume III, Vital Statistics Part 1, Table 19; under-five populations by sex are from the same source. Under-5 populations by sex are from the IPUMS 1900 full count data.[104] US areas in 1920 include 37 observations. These are comprised of rural and urban parts of MA, NY, PA, MD, IN, MI, OH, WI, and CA; the urban parts are exclusive of larger cities, which are included separately. The largest cities enter individually: Boston, Brooklyn, New York City, Philadephia, Pittsburgh, Chicago. Smaller cities are in urban aggreagates, as follows: other MA urban, urban CT, other urban New England; urban KS & MN; urban areas of the South; urban WA & OR. We also have: rural northern New England (ME, NH, VT), rural CT & RI, the rural parts of each of KS, MN, and VA; rural WA & OR; and the state of UT. For 1920, aggregates were formed to achieve a minimum under-5 population over 49,000 in order to reduce the role of random variation in sex-ratio data.

For 1920, infant mortality rates are calculated from on births and infant deaths for 1915–1919, taken from annual reports of birth statistics for the BRA.[105] The 1920 US census data refer to population as of January 1, 1920 so we take the simple averages (of births and of infant deaths) for the 5 years from 1915 to 1919.

Under-five populations by sex are from the IPUMS 1920 full count data.[106]

US areas in 1930 include 66 observations. Aggregates were formed to achieve a minimum under-5 population over 49,000 in order to reduce the role of random variation in sex-ratio data. These are comprised of rural and urban parts of New Jersey, New York, Pennsylvania, Illinois, Indianna, Michigan, Ohio, Wisconsin, Iowa, Missouri, Washington, and California; the urban parts are exclusive of larger cities, which are included separately. The largest cities were entered individually: New York City, Chicago, Detroit, Philadelphia, Los Angeles, Cleveland, Boston, Pittsburgh, St Louis. Smaller cities were grouped to varying degrees, as follows: Minneapolis & St Paul; San Francisco & Oakland; Baltimore & Washington DC, and other southern cities (New Orleans, Louisville, Atlanta, Memphis, Nashville). Cities smaller than those above are included in various urban aggregates, as follows: urban Massachusetts excluding Boston; urban New England excluding Massachusetts; West North Central urban (excluding Iowa and Missouri, included above); South Atlantic urban; other urban South

[104]Steven Ruggles, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. Minneapolis, MN: IPUMS, 2021.

[105]US Bureau of the Census, *Birth statistics for the registration area of the United States* : 1915, first annual report (Washington: GPO, 1917); 1916, second annual report (1918); and *Birth statistics for the birth registration area of the United States* 1917, third annual report (1919); 1918, fourth annual report (1920); 1919, fifth annual report (1921). These are available online at HathiTrust)

[106]Steven Ruggles, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. Minneapolis, MN: IPUMS, 2021.

(urban areas of states in the East South Central and West South Central census Divisions, exclusive of cities mentioned above). For 1930, we also have the rural parts of the states of Kansas, Minnesota, Nebraska, North Dakota, Virginia, Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Kentucky, Tennessee, and West Virginia.[107] Rural aggregates (for under-five populations over 49,000) include northern New England rural (ME, VT, NH), southern New England rural (CT, MA, RI), and rural Maryland & Delaware. With very small urban populations, we aggregated the smaller states Idaho & Utah, and Montana & Wyoming. Finally, for each of Colorado, New Mexico, and Oregon we use the entire state, because the urban portions fell well below our 49000 population-size threshold.

The 1930 data for California, Colorado, and New Mexico refer to total populations (white and nonwhite). Colorado births and infants deaths are not presented by race in 1930. For the other states, total populations are used because the 1930 census (unlike other censuses) classified persons deemed "Mexican" as non-white.[108]

For 1930, infant mortality rates are calculated from births and infant deaths for 1925–1930, taken from annual reports of birth statistics for the BRA.[109] The 1930 US census data refer to the population as of April 15, 1930; for an appropriate average IMR, we take weighted averages (of births and of infant deaths) across the 6 years 1925-1930; 1925 is weighted 260/365 of one-fifth, 1930 is weighted 105/365 of one-fifth, and the other 4 years each weighted one-fifth (thus we treat April 15 as 105 days through the year). Under-five populations by sex are from the IPUMS 1930 full count data.[110]

For the US in 1940 we use state-level data for the white population; setting a minimum under-five population of 25,000 we have 46 observations (Nevada, Delaware, and Wyoming being too small).

Under-five (white) populations by sex for 1940 are the simple average of the values from the published census and the IPUMS full count sample. The published census values are taken from the 1940 census, *Population Volume II, Characteristics of the Population, Sex, Age, Race, ...* That volume is presented in seven Parts; Table 7 for each state has data for the under-five white population by sex, DC's data appears in DC's Table 3.

Infant mortality rates for 1935–40 are taken from Linder & Grove (1947: Table 28 (pp. 578–605)). The 1940 , Characteristics of the Populationcensus date was April 1, so for the 5-year

[107]The urban parts of these states fell below our 49,000 population threshold, so they are included in urban aggregates (described above).

[108]See e.g. the 1940 Census (1943), *Population Volume 2, Characteristics of the population ..., Part 1: United States Summary ...* , p. 3). The 1940 census includes various corrected counts for the 1930 census, with "Mexicans" classified as "white" as in the census years other than 1930.

[109]US Bureau of the Census, *Birth, stillbirth, and infant mortality statistics for the birth registration area of the United States* 1925, eleventh annual report, part 1 (Washington: GPO, 1927); 1926, twelfth annual report, part 1 (1929); 1927, thirteeth annual report, part 1 (1930); 1928, fourteenth annual report (1930); 1929, fifteenth annual report (1932); 1930, sixteenth annual report (1934). These are available online at HathiTrust)

[110]Steven Ruggles, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. Minneapolis, MN: IPUMS, 2021.

average IMR we weight 1940 by 1/4 of one-fifth, 1935 by 3/4 of one-fifth, and the years 1936-39 by one-fifth each.

Sources for other Figures (4-9)

See text for sources for Figures 4,5,6,7,8, and 9, which draw on data detailed above or below.

Sources for Childhood Sex Ratios in the US

We draw on the decennial US censuses for under-five populations by sex, with two broad sources: published US census volumes and IPUMS "full count data"[111] IPUMS full count data are available for decennially 1850-1880 and 1900-1940. For these years, we average the census volume and the IPUMS full count values of under-5 populations by sex, taking each as a plausible tally of the underling census manuscripts.

Under-5 populations by sex for 1850, 1860, and 1870, for the US and for states, are reported in the *Ninth Census – Volume II. The Vital Statistics of the United States*: Tables XXIII (all races), XXVI (whites), and XXIX (Blacks).[112]
Under-five populations by sex for 1880 are reported in *Statistics of the Population of the United States at the Tenth Census (June 1, 1880)*.[113] National totals (white and nonwhite) are reported for single years of age in Table XX.[114] Table XXI reports state totals for these data.[115]

IPUMS "full count samples"[116] are available decennially for 1850–1880, for non-slave populations, and decennially from 1900–1940 (the 1890 census manuscripts have not survived).

[111] Ruggles et al (2021).
[112] US Census Bureau 1872, pp. 563, 575, 610, 619, 649, 658. "Race" categories follow usage in the source.
[113] The US Census website refers to this volume as "1880 Census: Volume 1. Statistics of the Population of the United States".
[114] Table XX. Population of the United States, by specified age, sex, race, ... 1880; pp. 48-49
[115] Table XXI. Population, by specified age, sex, race, ... 1880; pp. 552-645
[116] Ruggles et al (2021).

Data References

Backer, Julie E. (1961), *Dødeligheten og dens Årsaker i Norge 1856–1955; Trend of Mortality and Causes of Death in Norway 1856–1955.* Samfunnsøkonomiske studier nr. 10, Statistisk Sentralbyrå, Oslo. Available in Statistics Norway's Digitized publications (accessed 2021-10-05).

Caldwell, J. C. (1987). Chapter 2, "Population". In Vamplew (1987:32–41).

Galloway, Patrick R. (2007). "Galloway Prussia Database 1861 to 1914", www.patrickrgalloway.com.

Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org.

Linder F. E. & Grove, R. D. (1947). *Vital Statistics Rates in the United States 1900-1940* (Washington: GPO). Available online from the NBER.

McDonald P., Ruzicka, L. & Pyne, P. (1987). Chapter 3, "Marriage, fertility and mortality". In Vamplew (1987:42-61).

Mitchell B. R. & Deane, P. (1962). *Abstract of British Historical Statistics.* Cambridge, England: Cambridge University Press.

Palgrave Macmillan (Ed.) (2013). *International Historical Statistics* (eBook). London, England: Palgrave MacMillan Ltd. DOI 10.1057/978-1-137-30568-8.

Ruggles, S., et al. (2021). IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D014.V3.0

US Bureau of the Census (1975). *Historical statistics of the United States, colonial times to 1970, Bicentennial edition, part 1.* Washington, D.C.: US Government Printing Office. Available online at the United States Census Bureau.

US Census Bureau (1872). *Ninth Census – Volume II. The Vital Statistics of the United States,* (Washington: GPO).

US Census Bureau (1883). *Statistics of the Population of the United States at the Tenth Census (June 1, 1880)* (Washington: GPO).[117]

US Department of Health, Education, and Welfare (1954). History and organization of the vital statistics system. Chapter 1 (pp. 2–19) of *Vital Statistics of the United States 1950 Volume 1.* Washington D.C.: United States Government Printing Office.

US Department of Health, Education, and Welfare (1996). *Vital Statistics of the United States 1992, Volume II – Mortality.*

Vamplew, W., Editor (1987), *Australians – Historical Statistics.* Broadway, New South Wales: Fairfax, Syme & Weldon. Available online from the Academy of the Social Sciences in Australia.

[117] The US Census website refers to this volume as "1880 Census: Volume 1. Statistics of the Population of the United States".