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

Jesse McDevitt-Irwin and James R. Irwin*

09 September, 2022

**Abstract**

Basic facts of infant mortality in the 19th-century US remain to be established, due to a lack of vital statistics. We present new estimates of infant mortality for US whites 1850–1920, using childhood sex ratios. Because males are biologically more vulnerable than females to infant mortality, high rates of infant death tend to be reflected in female-skewed sex ratios. We offer a method for estimating infant mortality from childhood sex ratios, based on historical data from Europe and the US. Applying this method to 19th-century US census data, we find that infant mortality for US whites was just over 80 deaths per 1000 births in the period 1850–1880. Less than one-half of existing estimates from model life tables, our results suggest a major revision to prevailing views of US infant mortality in the 19th century. Sex-ratio evidence further suggests that racial inequality under slavery was far worse than previously thought, with Black infant mortality circa 1850 more than four times that of whites.

*Work in progress. Please do not cite. Comments and suggestions welcome.*

*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. This current version is a narrowed-down version of our older manuscript which was circulated under the title "Infant Sex Ratios and Infant Mortality in North America: 1850-1920." For helpful comments on this paper and previous versions, we thank Doug Almond, George Alter, Vincent Bagilet, Lou Cain, Joel E. Cohen, Kris Inwood, Claire Palandri, Samuel Preston, Jeffrey Shrader, and Rodrigo Soares. Data work and analysis was done in *R*. Citations for packages can be found in section 11.

# 1 Introduction

Infant mortality is a key indicator of population health and living standards, especially historically, when differences in infant mortality across populations were far greater than today.[1] Across 19th-century Europe the level of infant mortality was often the dominant component of living standards. For example, in 1850, nearly 1-in-3 infants born in Germany died before their first birthday, while in neighboring Norway that value was less than 1-in-10 (Mitchell 1998: series A7). Turning to the 19th-C US, what can we say about infant mortality – and therefore living standards?

Unfortunately, the level of infant mortality in the 19th-century US remains unknown, due to a lack of vital statistics data. Without records of births or infant deaths for most of the country, conjectures from model life tables are the closest we come to estimates of infant mortality for the 19th-century US as a whole.[2] Until we can characterize the level of infant mortality in the 19th-century US, we can't answer basic questions about contemporary living standards.

We offer a new method for characterizing broad patterns of infant mortality, using childhood sex ratios. Because of the biological survival advantage of infant females, high rates of infant mortality tend to skew the surviving population towards girls. Assembling historical vital statistics and census data, we document a striking relationship between childhood sex ratios and infant mortality. Taking childhood sex ratios from historical US census data, we use this relationship to infer contemporary infant mortality. We find that infant mortality among US whites was just over 80 deaths per 1000 in the period 1850–1880. Our new estimates are less than half of those implied by model life tables, suggesting a major revision to existing characterizations of infant mortality for the 19th-century US. On our evidence, US whites were among the healthiest populations of the 19th century. Sex-ratio evidence corroborates prevailing views of very high Black infant mortality, implying that racial health inequality under slavery was far worse than previously thought. By our new estimates, circa 1850, Black infant mortality was more than 4 times that of US whites.

The rest of this paper goes as follows: Section 2 reviews existing estimates of US infant mortality within a broader historical context; Section 3 presents the demography and biology

[1] In absolute terms, historical variation in infant mortality dwarfs that of the past two decades. For example, infant mortality rates ranged from 100 to 300 (per 1000) in 19th-century Europe (see data appendix, below). Today over one-third of the world population lives in places with infant mortality rates below 10, 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)).

[2] Most notable are Haines (1979, 1998) and Hacker (2010), studies which seek to characterize 19th-century US mortality across the lifespan, not to offer estimates of infant mortality.

underlying the relationship of infant mortality and childhood sex ratios; Section 4 uses historical data from Europe and the US to characterize the empirical relationship between childhood sex ratios and infant mortality; Section 5 presents our new findings on 19th-century US infant mortality; Section 6 demonstrates the robustness of our results; Sections 7 and 8 close the paper. Section 7 elaborates on our results with reference to known patterns of mortality in contemporary Europe, pointing toward potential explanations for our finding of relatively low infant mortality among 19th-century US whites. Section 8 concludes by summarizing our argument and some broader implications.

## 2 Historical Background

The basic facts of white infant mortality in the United States since the mid-19th century may appear to be reasonably complete.[3] 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.[4] Figure 1 plots the *HSUS* series against the backdrop of IMR data available for a cross-section of European polities from 1840–1990. The *HSUS* series features high rates on 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 resumed, 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: what stands out is a general pattern of massive improvement in infant mortality since the late 19th century.

The sustained decline in the US white IMR series since 1880 may appear to be just another facet of the widely studied "mortality revolution" (Easterlin 1996), which has seen life expectancies soar and mortality rates plummet across the globe in the twentieth century.[5] Narratives of a mortality revolution constitute the dominant paradigm of historical

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

[4]HSUS Series Ab921. The underlying data come from Haines (1979), discussed below. Previous editions of HSUS, produced and published by the US Census Bureau, presented US infant mortality rates starting with 1915, based on vital statistics data. See below, for discussion of the IMR values at census benchmarks from 1850 to 1910, which appear in the current (2006) edition of HSUS; this edition was "prepared by the academic community" and published by Cambridge University Press (HSUS 2006: Appendix 3; "Editions and Copyright"; see also "Editor's Preface").

[5]Since the 19th century, global average life expectancy at birth has risen from near 30 to over 70 (Wiley 2005: table 1), and under-5 mortality has plummeted from roughly 1-in-2 (Hill 1995) to nearly 1-in-30 (UNICEF). See also Riley (2005a) and Costa (2015), who use the term "health transition" to refer to the widespread huge improvements in human health.

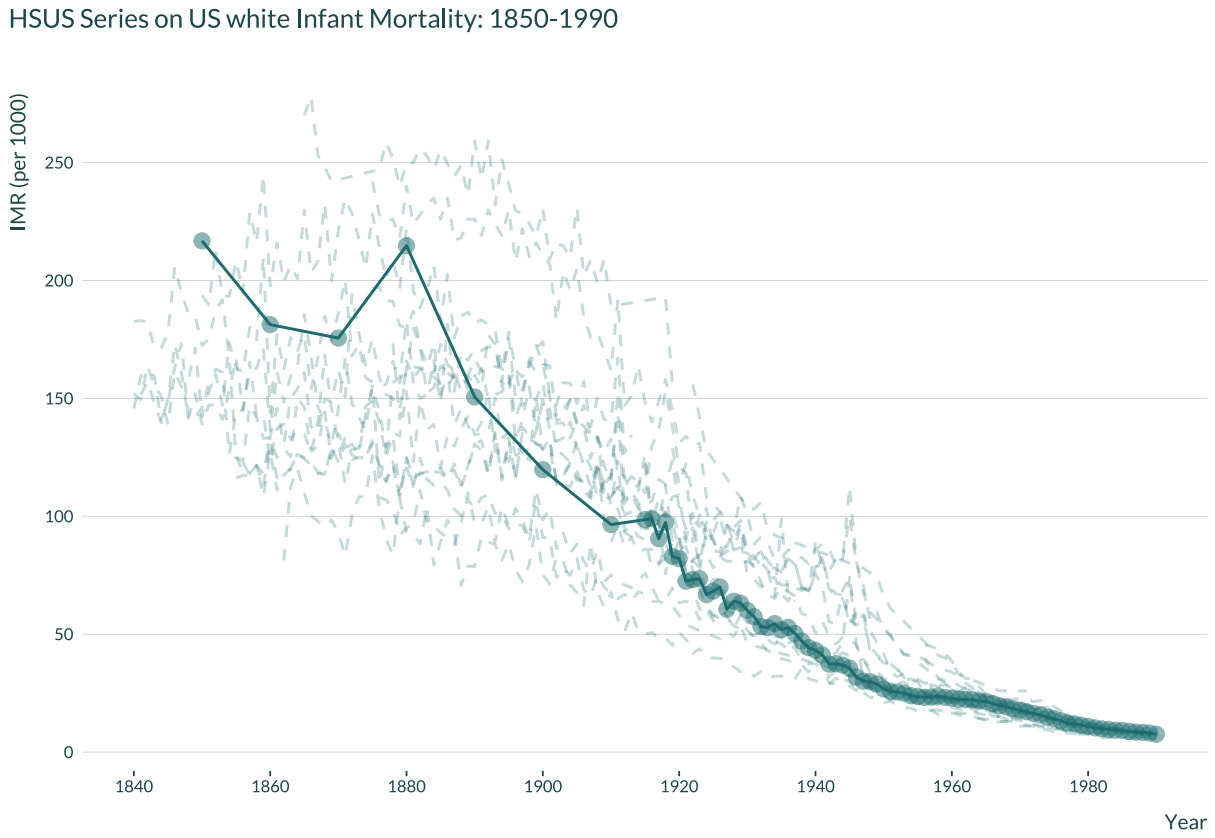


Figure 1: Infant Mortality Rates, 1850-1990. The solid line is US whites; the dashed lines are various European populations. Sources: HSUS series Ab921; European data from Mitchell (1998).

research on mortality, including a wealth of studies investigating the emergence of the very low mortality regimes enjoyed in the 'developed' world.[6] 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 shared historical knowledge is evident in the opening line of Anderson et al. (2022:126): "Since the mid 19th century, mortality rates in the Western world have plummeted and life expectancy has risen dramatically."[7] Central also to the paradigm is a broad historical explanation for the collapse of mortality: that the scientific and industrial revolutions were "transmuted" into the mortality revolution (Caldwell 2006). This collapse in mortality thus plays a central role in the broader narrative of 'progress' surrounding 19th and 20th century modernization and industrialization.[8] Within this context, the tremendous decline of infant mortality seen in the *HSUS* series may seem unremarkable.[9]

However, the *level* of infant mortality presented in *HSUS* for the 19th-century US is puzzling in light of infant mortality evidence from both sides of the Atlantic. Infant mortality around 200 would rank US whites among the unhealthiest of contemporary European populations, with double the infant mortality of Norway (IMR of 100), much worse than England or France (IMR 150–160), and nearly as bad as Germany and Austria (IMR $>$ 200).[10] Such poor health is difficult to reconcile with the established fact that US whites were among the tallest people in the world during the 19th century (Fogel et al. 1983:463).[11] In addition, the US was predominantly rural during the 19th century, when there was a substantial "urban penalty" (Kearns 1988) in mortality (e.g., Davis 1973:100–105; Williamson 1982; Haines 2001; Cain and Hong 2009).[12] In the 19th century, US whites were a tall, mainly rural

[6] Among many contributions addressing the mortality revolution, 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).

[7] As befits facts that are common knowledge, the statement is without references or citations.

[8] This narrative of progress is the subject of the very work which coined the term 'mortality revolution': Easterlin's (1996) *Growth Triumphant*.

[9] 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 Mort Rev paradigm even better than does the *HSUS* series.

[10] IMRs varied annually in all these countries; here we are presenting broad patterns.

[11] Even after a decline in heights from (birth-years) 1830–1860 – the oft-discussed 'antebellum puzzle' (see, e.g., Margo and Steckel 1983; Komlos 1996) – US whites remained one of the tallest populations in the world. 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).

[12] For example, in 1890 England, urban infant mortality was about 220, while rural was just under 100 (Woods, Watterson, and Woodward 1988). If the US had a similar urban penalty, then the HSUS values for

population which drew millions of voluntary immigrants from Europe: infant mortality of 200 from 1850–1880 would constitute a serious puzzle.

Attention to the slim body of direct evidence on infant or childhood mortality in the 19th-century US raises further questions about the high level of infant mortality shown in the HSUS series. Most notably, there are (we have) highly credible vital statistics data for the state of Massachusetts from circa 1860. For that state, we see an infant mortality rate of below 160 (on average) in the period 1860-80 (or 1860-1900) – a remarkable 45 points below the level of mortality suggested by HSUS Series Ab921 for the US white population 1850-1880. But vital statistics data from 1890 and 1900 clearly identify Massachusetts as a high mortality state, as expected from its very high level of urbanization.[13] As noted above, there was a substantial urban penalty for infant mortality in the 19th century, and US vital statistics show that the US was no exception to this rule (Condran and Crimmins 1980). 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 country suggests that infant mortality was closer to 100 than 200 in the mid 19th century. Lynch et al. (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 the same direction. Using census data on maternal recall to estimate child mortality, Haines (1977: table 4) estimates rural under-5 mortality of 18–19% and urban 25–26% (table 4). Those levels of child mortality suggest rates of rural and urban infant mortality around 110 and 150 respectively.[14] More broadly, Haines and Preston (1991: table 2.2, sources therein) report a series of IMR estimates for the 19th-century US which cast doubt on the *HSUS* series, with IMR ranging from 135 in New York state to 160–175 in Philadelphia and above 200 in New York city. The US was over 70% rural in 1880, making it highly doubtful that US whites could have had as high infant mortality *as a whole* as the largest cities in the country. Indeed, once death data become available around the turn of the century, it is

1850–1880 imply that *rural* US white infant mortality was greater than the infant mortality of England as a whole (authors' calculations).

[13]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).

[14]Using the Coale and Demeny (1983:46-49) West model life tables , the average of levels 10 and 11 for urban, and 14 and 15 for rural. Our suggested IMR values for upstate NY are conjectural, because there is no fixed relationship between infant mortality and mortality at ages one to four, as we discuss more fully below. For example, using the East model life table would imply higher infant mortality, and the North lower.

clear that the highly urbanized Northeast region has higher infant mortality than the rest of the country.[15] All of this evidence points to US whites as a whole having infant mortality closer to 100 than to 200 from 1850–1880. The high level of 19th-century infant mortality seen in the *HSUS* series thus appears at odds with known patterns of population health in both Europe and the US.

However, the *HSUS* values for the 19th century should be seen as conjectural. Unlike those from 1900 onward, the IMR values for the 19th century are not based on any contemporary evidence of births or infant deaths. In terms of sources and methods, the *HSUS* series is a composite (Haines 2006), as shown in Figure 2. From 1933 onward, the values are from nationwide vital statistics. From 1915–1932, the values are from vital statistics for 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). The values for 1900 and 1910 are indirect estimates of infant mortality, based on census data of maternal recall of births and deaths.[16] The 19th-century values were constructed without evidence on infants' births or deaths, simply for a lack of credible data. The requisite data for standard methods of estimating IMR – directly or indirectly – are not available for the nineteenth-century US. Therefore, Haines (1979; 1998) fits a model life-table system to census data on mortality at older ages.[17] Once estimated, these life table include the level of infant mortality for each census year. The *HSUS* IMR values are thus extrapolations from older-age mortality. The level of infant mortality is determined by the specific age-pattern of mortality assumed in the model life table.[18]

However, infant mortality is not implied by mortality at older ages.[19] As forcefully argued by Woods (1993:217), infant mortality, child and adult mortality are each "indispensable" indices, "since 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

[15] 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).

[16] Though labelled 1900 and 1910, these estimates "are from approximately 1895 and 1904, respectively" (HSUS 2006:note2), due to the nature of the estimation methods (Preston and Haines 1991:74, Haines 1998:154).

[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 genealogical estimates of adult mortality rather than census mortality data. The implied infant mortality values are broadly comparable to Haines, as shown in Figure 2.

[19] Hacker (2010:76) makes this exact point within the context of the 19th-century US, noting the "weakness" of basing entire life tables on estimates of adult mortality. Moreover, in Table 6 he illustrates the wide range of infant mortality possible when extrapolating from adult mortality.

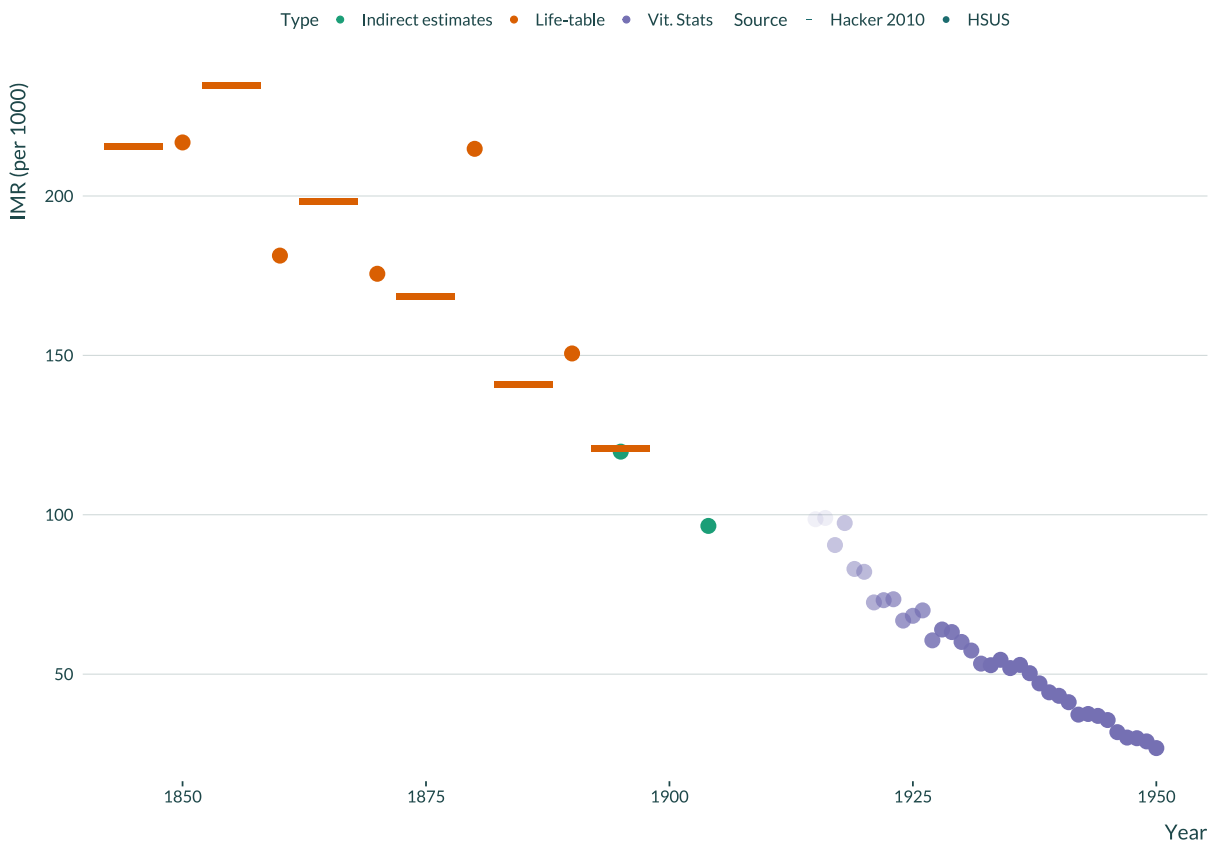


Figure 2: US white Infant Mortality by Year, colored by the source of data. 'Life-table' refers to estimates that are extrapolations from older-age mortality using life tables. 'Indirect estimates' refers to the 'children-ever-born / children-surviving' method (Preston and Haines 1991). Vital statistics points are shaded according to the completeness of coverage, going from 1/3 in 1915 to 100% in 1933. The circles are from *HSUS* (2006). The dashes are from Hacker (2010: table 8) and are decadal estimates.

age 5–20 mortality declined by half from 1840–1880, yet infant mortality remained unchanged (HMD data).[20] In such a case, extrapolating from age 5–20 mortality would produce severe overestimates of infant mortality in preceding periods. More generally, European life tables (based on vital statistics data) include a wide range of infant mortality rates relative to mortality at older ages. Figure 2 plots infant mortality rates against age 5-20 mortality rates from Europe 1850–1920; the shaded area shows the range of age 5-20 mortality rates in Haines's life table that produced the HSUS IMR estimates (1998: appendix A). Given this range of age 5–20 mortality, the infant mortality rate could be anywhere from below 80 to above 200.

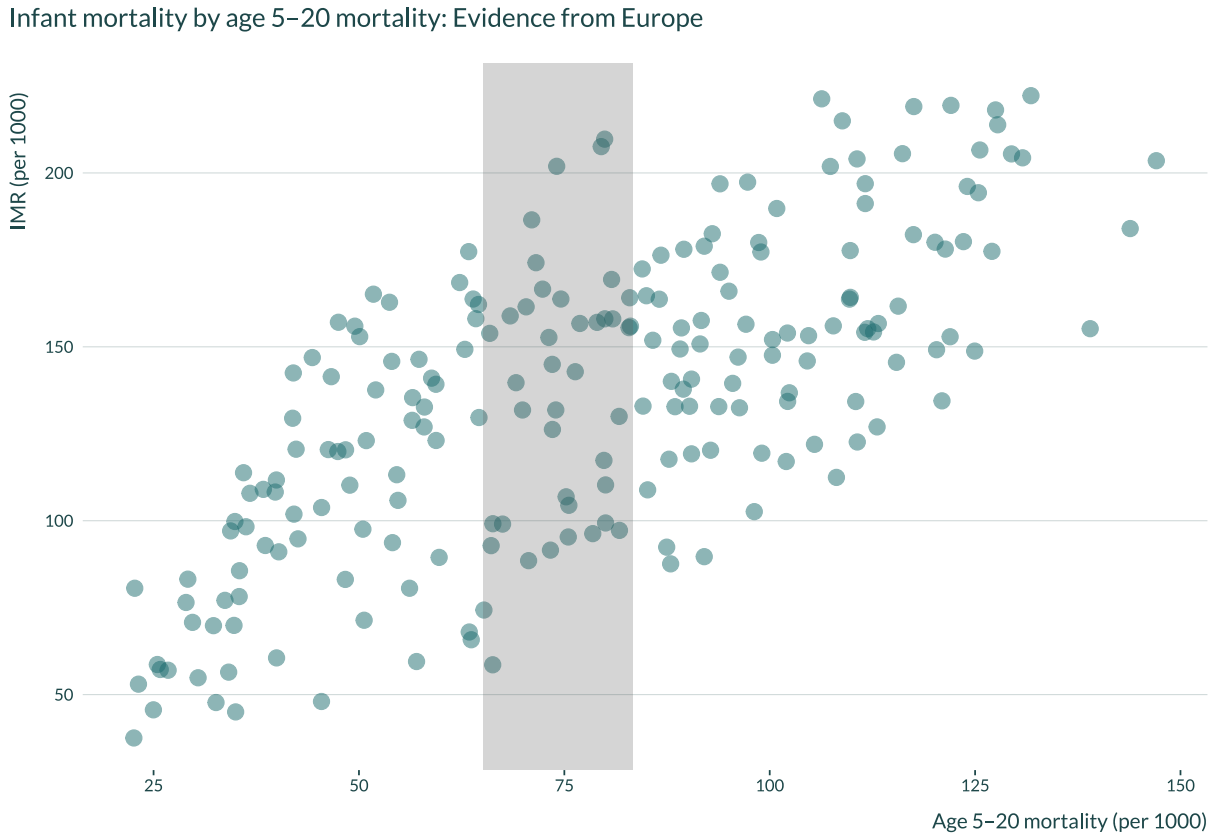


Figure 3: Infant mortality by age 5–20 mortality. Source: *HMD*. The shaded area is the range of estimates of age 5–20 mortality for US whites 1850–1880, as given by Haines (1998:156–165).

When uncertainty over the level of older-age mortality is combined with a lack of data on the age-pattern of mortality, we get a staggering range of possible IMR for 19th-century US whites. For example, if we take Hacker's (2010: table 5) estimate of US 1850s male

[20]The English example is well-studied: see e.g., CITES. Conversely, in Sweden over the same period age 5–20 mortality remained unchanged while infant mortality declined by over a third (HMD data).

life expectancy at age 20 (*e20*), 38.4 years, and fit a Coale-Demeny East life table, we get male infant mortality of 281. If we take Ferrie's (1996: table A1) estimate of *e20* in 1850, 45.4 years, and apply a Coale-Demeny North life table, we get male infant mortality of 90.[21] Without direct evidence from vital statistics, we simply don't know where in this range US infant mortality actually fell within this range.

Further empirical research on 19th-century US infant mortality is thus "sorely needed," but is limited by a lack of data (Hacker 2010:76). We are able to provide new evidence on 19th-century US infant mortality. Building off well-known facts of biology and demography, we have devised a new method for characterizing broad patterns of infant mortality, using childhood sex ratios. Our evidence places 19th-century US whites' infant mortality as substantially lower than existing estimates, and as one of the healthiest populations of its time.

# 3 Childhood Sex Ratios Reveal Infant Mortality[22]

It has long been known that biologically, girls are less vulnerable than boys to infant mortality.[23] The corollary which we highlight, and build from, is that high rates of infant mortality tend to skew childhood sex ratios toward females, unless girls' biological advantage is offset by sex discrimination.[24] This pattern is apparent in both historical populations and familiar model life tables. For example, in 1900 infant mortality in Austria was above 200, and there were roughly the same number 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

[21] The first IMR comes from East level 8 (Coale and Demeny 1983:273); the second comes from North level 16 (Coale and Demeny 1983:163).

[22] This section draws on our working paper,"Childhood Sex Ratios Reveal Infant Mortality" (McDevitt-Irwin and Irwin 2022).

[23] 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 an authoritative review see Waldron (1998:64–83).

[24] The female survival advantage can be nullified by extreme sex discrimination, making male-skewed sex ratios in the presence of high infant mortality a clear sign of 'missing women'. See e.g., Beltrán Tapia and Raftakis (2021), especially Figures 2 & 3. Existing evidence on child mortality from the 19th-century US show a clear female survival advantage (Haines 1977: table 7; Kunze 1979: table 14; Lynch et al. 1984: table 4), so 'missing girls' should not be a concern for our analysis of the US. However, the possible effects of sex discrimination have to be kept in mind when using childhood sex ratios to make inferences about infant mortality. See (McDevitt-Irwin and Irwin 2022) for further discussion.

typical of the sex ratio at birth in healthy populations (Maconochie and Roman 1997; Grech, Savona-Ventura, and Vassallo-Agius 2002). A similar pattern can be seen across Europe, as we document in Figure 4. In familiar life tables, the pattern is also present: e.g. in the Coale and Demeny West model, moving from level 11 to level 22 infant mortality plummets from 159 to 27 (per 1000) and the sex ratio among survivors to age five ($_5l_0$) shifts 2.14 percentage points away from girls (Coale and Demeny 1983:47,52).

A simple model illustrates how the sex ratio among surviving children is determined by the combination of the level of infant mortality, the degree of excess male mortality, and the sex ratio at birth. Define the childhood sex ratio as the (log) sex ratio of the hypothetical population of survivors to age 1, denoted $SR1$.[25] Let $b^i$ and $q^i$ be the number of births and infant mortality rate of sex $i$;[26] the SR1 is given by $SR1 = ln(\frac{b^f \cdot (1-q^f)}{b^m \cdot (1-q^m)})$. Defining the sex ratio at birth as $SRB \equiv ln(\frac{b^f}{b_m})$, a few steps of algebra take us to the following representation:

$$(1) \quad SR1 = SRB + [ln(1-q^f) - ln(1-q^m)].$$

The SR1 is determined by two, additively separable terms: (1) the sex ratio at birth, and (2) the relative survival of boys and girls. As infant mortality approaches zero, so does the second term, and the childhood sex ratio approaches the sex ratio at birth.

We can further simply this expression by taking a Taylor series approximation ($ln(1+x) \approx x$), and defining $q$ as overall infant mortality and $\mu = \frac{q^m - q^f}{q}$ as excess male mortality. We obtain:

$$(2) \quad SR1 \approx SRB + \mu \cdot q$$

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, $\mu$, the more that infant mortality skews the sex ratio among survivors. Importantly, this effect is proportional to the level of infant mortality,[27] so the effect will be negligible for populations with low infant mortality (e.g. rates below 10). 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%

[25] We model the sex ratio at age 1, as age 1–4 mortality is unlikely to affect the childhood sex ratio as it is much less male skewed than infant mortality (Hill and Upchurch 1995).

[26] Note that taking life-table notation we would be working with $l_1^i$.

[27] In fact the effect is slightly more than proportional to q, because of our Taylor series approximations.

more boys than girls – it would take infant mortality of 150–300 to drive the sex ratio of the surviving population to parity.

Of course, we would not expect a population with such high infant mortality to have a healthy sex ratio at birth, as a growing body of work demonstrates that insults to maternal wellbeing push the sex ratio at birth towards females (e.g., Almond and Edlund 2007; Fukuda et al. 1998; Catalano 2003).[28] 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 past their first year. The direct effect of infant mortality on childhood sex ratios will, in most cases, be reinforced by a female-tilted sex ratio at birth, as infant mortality and maternal health are closely linked (e.g., Kramer 1987).[29]

Therefore, any structural interpretation of a regression of childhood sex ratios on infant mortality – equation (2) – would have maternal health as a latent variable. As described by (Goldberger 1973:5), the estimated regression coefficients would correspond to "mixtures" of the underlying structural parameters. In our case, the estimated slope coefficient would not only capture the direct effect of infant mortality on sex ratios – i.e. 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 pattern 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.[30] In practice, it means we can expect different CSR from the same level of IMR when populations differ in the terms of the nature and causes of IMR. 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 ratio reflect, and therefore can predict, infant mortality is an empirical question, one which we answer using historical data from vital statistics.

[28]The apparent mechanism is maternal stress hormones, which increase the probability of miscarriages, which are disproportionately male (James and Grech 2017:51). 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 2020).

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

[30]Consider, for example, breastfeeding. It largely unrelated to maternal health, but plays a huge role in infant mortality.

# 4 Empirical Approach

In order to characterize the empirical relationship of childhood sex ratios and infant mortality, we assemble data from the US and Europe, mostly from the mid-19th century onward.[31] We calculate the under-5 sex ratios, children aged 0–4, and pair this to the rolling mean of infant mortality.[32] We use the under-5 sex ratio for several reasons: it is widely available in published census data, pooling the under-5 population increases the sample sizes, and pooling across ages reduces the impact of sex-biased age heaping. We end our series in the early 1960s; by then infant mortality in our sample populations was low enough that patterns of sex-ratio variation were largely independent of infant mortality, and ultrasound, which spread in the 1970s (Campbell 2013), was not yet a factor in sex-ratio patterns. This leaves about 200 observations for Europe. For the period of 1900 to 1930, we dis-aggregate the US data by state and rural-urban, leaving us with about 140 observations from the US.[33]

Taking these data, we characterize the relationship between infant mortality and childhood sex ratios. We use this relationship to infer 19th-century US infant mortality from observed under-5 sex ratios, which are available from the US census.[34]

In Section 3 we argued that childhood sex ratios should reflect infant mortality. Within our sample, their actual empirical correspondence is striking. In Figure 4 we plot under-5 sex ratios against infant mortality: high rates of infant mortality imply relatively more girls, and low rates relatively more boys.[35] 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.

[31] See section 12 regarding our sample. In brief, our non-US data cover: Sweden (1757–1960), Denmark (1840–1960), Belgium (1846–1960), England and Wales (1851–1961), the Netherlands (1859–1960), Scotland (1861–1960), New Zealand (1867–1961), Austria (1869–1961), Australia (1880–1961), Germany (1880–1961), Switzerland (1880–1960), Finland (1885–1960), Norway (1890–1960), France (1901–1954), Italy (1911–1961), and South Africa (1918–1921). For the US we have Massachusetts from 1860–1960, and then a growing number of states from 1900 onward.

[32] For the European data, we pair the under-5 sex ratio to the 5-year rolling mean, but for some of the US states we have as few as one year of infant mortality data.

[33] After 1930, state-level US variation in white infant mortality is too small to be of interest for our study.

[34] We calculate under-5 sex ratios for the US white population from 1850 onward, drawing on both published census data and the full-count PUMS. The 1890 Census records were lost in a fire, so there is no PUMS data available. Moreover, age reporting in the 1890 census was not consistent with practices in the other censuses. As discussed in 1900 and 1910 census reports, only the 1890 census asked for "age at nearest birthday" instead of "age at last birthday", which was used from 1850 to 1880, and from 1900 forward (page xlviii of Twelfth Census (1900), Census Reports Volume II, Population Part II, Washington: GPO, 1902). Therefore we exclude 1890 from our analysis.

[35] We note one slight outlier – Italy – which has much more male-skewed sex ratios than would be expected from infant mortality. We break down the Italian regions, and find that the male-skewed sex ratios are driven entirely by one region (Southern Italy), suggesting that Italy had 'missing girls', much as found for contemporary Greece (Beltrán Tapia and Raftakis 2021).

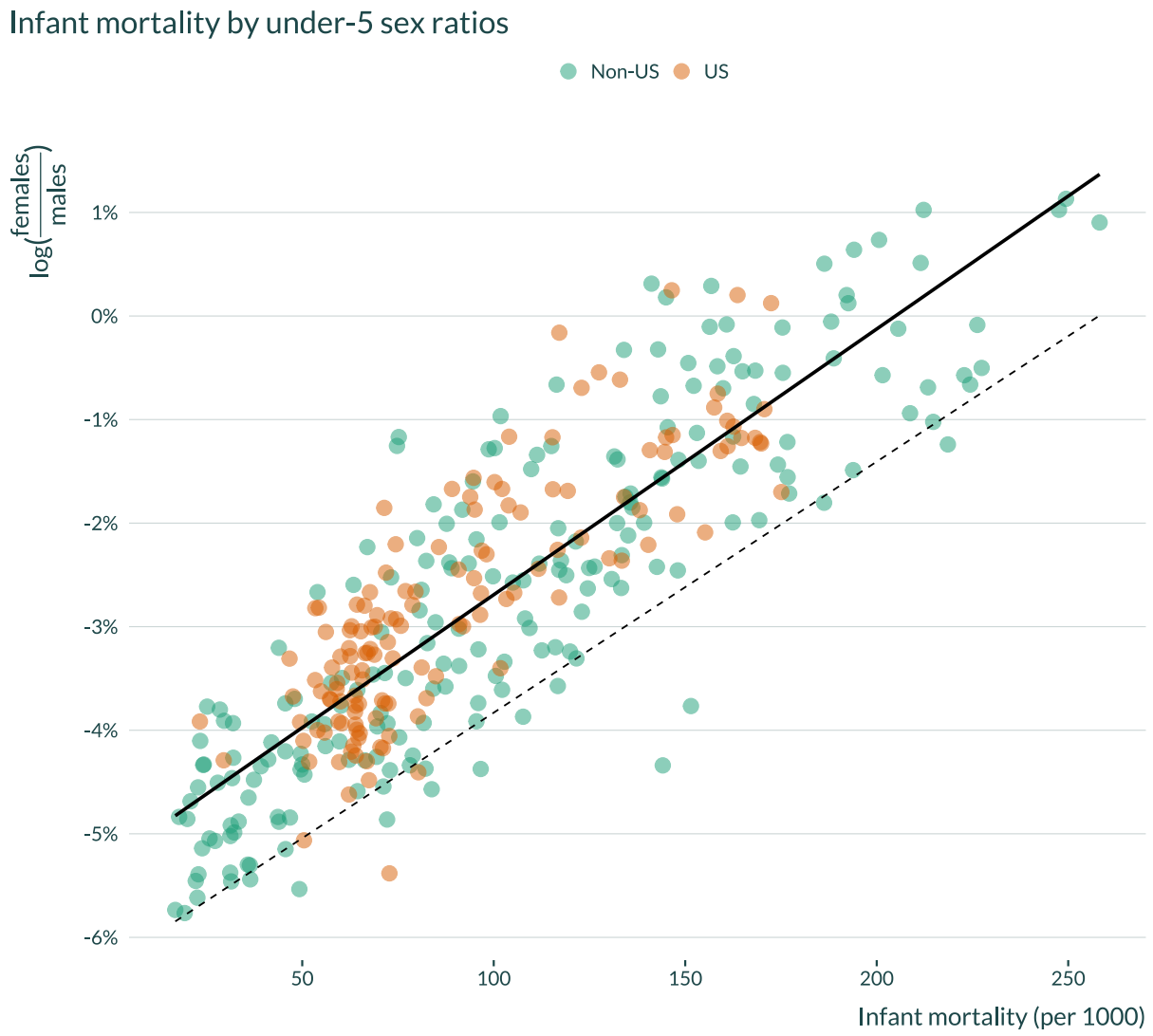


Figure 4: Infant mortality by under-5 sex ratios. The black line is the OLS regression of under-5 sex ratios on infant mortality. The dashed line is the 5th percentile, as estimated by quantile regression. Data mainly from Europe and US, see note X.

In order to characterize their relationship, we first regress the under-5 sex ratio ($SR$) on infant mortality ($IMR$). This equation, *R1*, is plotted in Figure 4:[36]

$$(R1) \quad \hat{SR}_i = -5.258 + 0.02566 \cdot IMR_i$$
$$(0.08158) \quad (7.920 \times 10^{-4})$$

These estimates predict that a population with zero infant mortality would have 5.4% more boys than girls, and that a 100 point increase in infant mortality (per 1000 births) would cause the under-five sex ratio to move 2.6 percentage points towards girls. These values closely correspond to those suggested by equation (1) above, with 5% more boys than girls being a healthy sex ratio at birth (Maconochie and Roman 1997; Grech, Savona-Ventura, and Vassallo-Agius 2002) and 25% being within the normal range of excess male mortality (Hill and Upchurch 1995; Alkema et al. 2014). With an eye toward predicting infant mortality from sex ratios, the *R1* results suggest that looking across populations, the IMR is predicted to be about 78 points lower in a population with a population that is 2% more male.

Our goal is to predict infant mortality from under-5 sex ratios, therefore we also run this regression in reverse, taking under-5 sex ratios as the explanatory variable.[37] We obtain:

$$(R2) \quad \hat{IMR}_i = 177.3 + 28.81 \cdot SR_i$$
$$(3.347) \quad (0.9597)$$

These results predict that a 1% change towards boys would be associated with a 29 point decrease in infant mortality, and that a population with equal number of boys and girls would have infant mortality of about 180.[38]

Our goal is to infer 19th-century US infant mortality from under-5 sex ratios. The simplest approach is to use *R2* to give point estimates of infant mortality given an under-5 sex ratio. As proof of concept, we apply this prediction method to Massachusetts, which was

[36] N = 351. Heteroskedasticity-robust standard errors reported in parentheses.

[37] Using both of these regressions helps us to address the effect of measurement error, which is a of particular concern for childhood sex ratios. 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 (1967:33), 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.

[38] Based on *R1*, sex-ratio parity is predicted with IMR of 205.

the only US state with vital statistics going back to the mid-19th century. We drop the Massachusetts data, re-estimate equation (3), and then predict IMR from under-5 sex ratios in Massachusetts. Plotted in Figure 5, we find a striking, if rough, correspondence between predicted infant mortality and the actual values. The Massachusetts example illustrates the promise of childhood sex ratio evidence for characterizing the approximate level of infant mortality in a population.

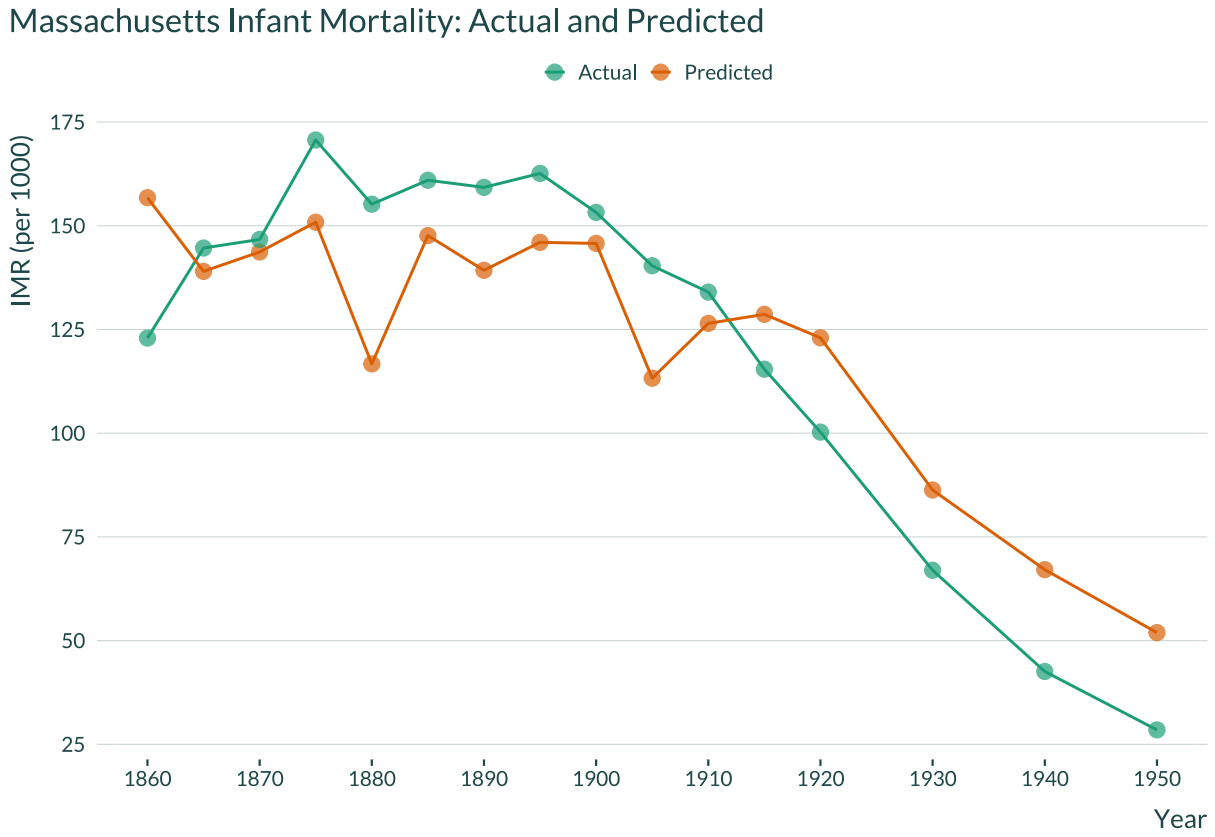


Figure 5: Out-of-sample prediction of infant mortality from childhood sex ratios: Massachusetts. Predicted values calculated from a regression of infant mortality on sex ratios, excluding Massachusetts data. IMR from *HSUS* series Ab928.

We systematize this process using out-of-sample testing. We drop the observations from one country, regress infant mortality on sex ratios in the remaining data, and then predict the infant mortality for the dropped observations.[39] The 80% prediction interval is roughly ±35 points, and the 2/3 interval ±25 points. The IMR predictions made from sex ratios are coarse, and are useful as interval estimates, as opposed to point estimates. Given this level

[39]We build off of the cross validation approach proposed by Butler and Rothman (1980); more recent literature refers to this type of technique as a 'jackknife' (Barber et al. 2021).

of uncertainty, childhood sex ratios should be able to distinguish between infant mortality of 100 vs 150 (like Norway vs England), but not 150 vs 160 (like England vs France).

Next, we use the logic of hypothesis testing to construct an upper bound on infant mortality, given an observed under-5 sex ratio. We characterize the conditional distribution of under-5 sex ratios on infant mortality, and then estimate the likelihood of observed sex ratios given some level of infant mortality. Using bivariate quantile regression, we estimate the $\tau$ quantile of the conditional distribution of under-5 sex ratios on infant mortality: $q_{SR|IMR}(\tau) = \alpha + \beta \cdot IMR$. Setting $\tau$ to the 5th percentile, we construct a one-sided hypothesis test of 95% significance. For an observed under-5 sex ratio of $SR_i$, we reject all infant mortality beyond the level which corresponds to this 5th percentile: i.e. reject $IMR > \bar{IMR}$, where $\bar{IMR} = \frac{SR_i - \alpha}{\beta}$.[40] Graphically, given an observed sex ratio, we reject all infant mortality to the right of the 5th percentile line plotted in Figure 4.

Thus we have two approaches for quantifying the uncertainty in our estimates. First, a prediction interval using out-of-sample testing; second, an upper bound approach using quantile regression. The first method is conceptually Bayesian, as we characterize the distribution of infant mortality (the unknown variable) conditional on sex ratios (the observed data). The second method is conceptually frequentist, as we characterize the likelihood of observed sex ratios conditional on hypothesized infant mortality. These two methods each contribute a different qualitative result. The hypothesis test enables us to test the plausibility of existing estimates of infant mortality, while the prediction interval suggests a new range of probable infant mortality given observed under-5 sex ratios.

# 5 Results

US under-5 sex ratios in the mid 19th century saw over 3% more boys than girls – simply inconsistent with the HSUS estimates that place US infant mortality in the period 1850 to 1880 at about 200 deaths per 1000 (see Figure 2). Such high infant mortality has 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 4 – 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.

[40] Our regression for the 5th percentile returns $\alpha = -6.254$ (0.1901) and $\beta = 0.02424$ (0.001689). Standard errors (Huber) in parentheses.

The observed under-5 sex ratios of US whites baldly contradict these implications, ranging from 3.1% more boys than girls for 1870 to nearly 3.5% for 1850. Put simply, an infant mortality rate on the order of 200 would have skewed the childhood sex ratio markedly toward girls. To accept the prevailing estimates of US white infant mortality is to claim that the mid-19th-century US deviated profoundly from the experiences of contemporary Europe, as well those of 20th-century American populations.

We plot our new estimates of IMR for US whites alongside the *HSUS* series in Figure 6. For the period of 1850–1880, our point estimates (from *R2*) of US infant mortality fall between 77 and 87 deaths per 1000, with a 2/3 prediction interval roughly spanning 55–105.[41] Applying the hypothesis test described above, at the 5% significance level we exclude average infant mortality greater than 125 for the period 1850–1880. 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 of 150) during the period. We fail to reject the hypothesis that the US had similar infant mortality to Scotland, where infant mortality was 120, though we find it unlikely (e.g. at the 10% level we could reject it). By our estimates, infant mortality increased modestly from 1880 to 1900, before beginning the well-documented 20th-century decline.[42] Our 20th-century estimates line up well with the values in the *HSUS* series, albeit with an upward bias.

# 6 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 4). In this sense, our qualitative result that US mortality was around 100 or less is very robust, and is not sensitive to restrictions of our sample (such as using only US data, or data from before 1900) or modifications to our empirical specification (such as allowing for non-linearity).[43]

However, one potential concern is the quality of the under-5 sex ratio data itself. Under-enumeration of young children is a common problem in historical censuses, and has been explored in the case of the 19th-century US (see, e.g., Coale and Zelnik 1963; the various works published in Volume 15 - Issue 4 of *Social Science History*; Hacker 2013). If under-enumeration was sex-biased, then it could affect the measured childhood sex ratios used

[41] An alternative method of estimating infant mortality would be to apply the slope from *R2* to sex-ratio differences and extrapolate from the US IMR of 1930. Under this method, we obtain estimates around 70.

[42] Even at this 1900 peak, our estimates place US white IMR as no worse than seen in relatively healthy contemporary European populations.

[43] See Figure 10 in the appendix for these results.

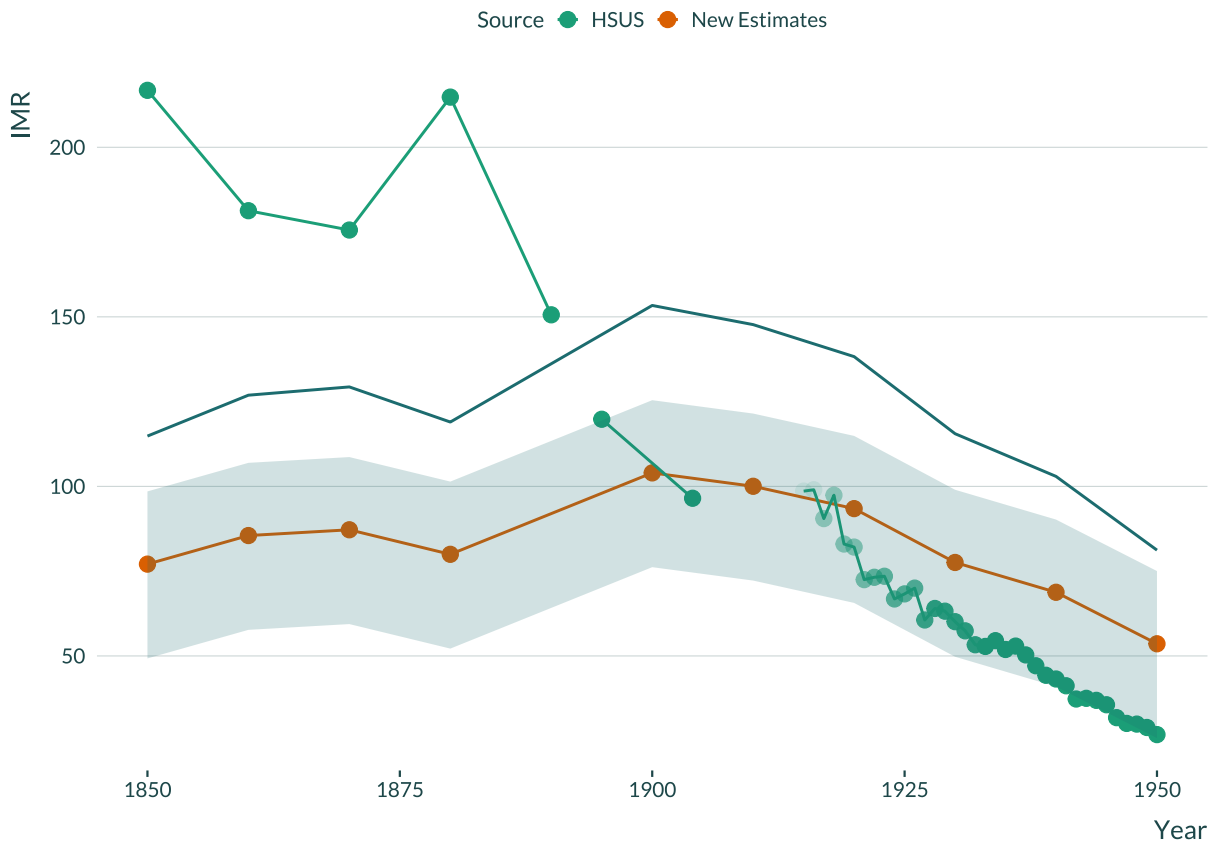


Figure 6: New Estimates of US white infant mortality: 1850–1950. Point estimates from equation *R2*. Shaded ribbon is the 2/3 prediction interval, calculated by out-of-sample testing. Upper bound calculated by quantile (5th percentile) regression of under-5 sex ratios on infant mortality. Existing estimates from HSUS series Ab921, broken into three segments according to source. The first segment comes from life tables, the second from indirect estimates, and the third from vital statistics. See Figure 2.

throughout this paper. More specifically, if under-enumeration patterns changed from the 19th century to the 20th, then the patterns observed in our sex-ratio data could reflect enumeration changes rather than actual population patterns.[44]

To test whether are results can be explained as artifacts of 19th-century enumeration problems, we compare under-five sex ratios in one census to the age 10–14 sex ratio in the census 10 years later – essentially following the cohort across the decade. We look at the nation as a whole, and only the US-born, so that immigration and inter-regional migration are not at play. The age 10–14 sex ratio promises to be a useful 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 an undercounting of infant girls in the 19th century is biasing our under-5 sex ratios toward boys – for a false impression of low infant mortality – then we should observe a relatively more female sex ratio among 10–14 year-olds ten years later.

However, attention to the 10–14 sex ratios only strengthens our results. In Figure 7 we plot under-5 sex ratios ('current') for each census year with the 10–14 sex ratios ('future') from the following census (10 years later).[45] The two measures of childhood sex ratios line up well, with similar levels and trends.[46] Comparing current and forward sex ratios in 1850 or 1870 suggests that any sex-biased undercounting was biasing under-5 sex ratios away from, not towards, boys. For 1850 and 1870, the forward CSR suggests a more male population of children; for 1860 the forward sex ratio is almost the same as the current. For 1900-1930, we do see some sign of relative underenumeration of infant girls, suggesting that the under-5 sex ratios were biased toward healthy-looking. But this only strengthens our argument that sex-ratio evidence suggests a deterioration of maternal-infant health between 1850 and 1900.

Our basic findings are thus corroborated by under-enumeration tests. Furthermore, the under-enumeration exercise allows us to see the pattern of sex ratios for 1890, a year excluded from our under-5 data because of data concerns (see footnote X). Future census data place 1890 childhood sex ratios as in between those of 1880 and 1900, but closer to the levels found in 1900, again reinforcing our finding that US white infant mortality was lower in the mid-19th century than in 1900. Similarly, we can use the 10–14 age group of the 1850

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

[45] Recall that we exclude 1890 because the enumeration of ages was not consistent with that in the other censuses, and because there are no PUMS data; so we do not have a future sex-ratio value for 1880, nor a current value for 1890.

[46] We do the same tests for many combinations of age groups (e.g. ages 0–5, 2–3, 1–4, 0–1) and find similar results.

census to make inferences about 1840 infant mortality. We find that infant mortality likely *even lower* in 1840 than in 1850.[47] This evidence corroborates the prevailing view that US population health deteriorated across the first half of the 19th century.[48]

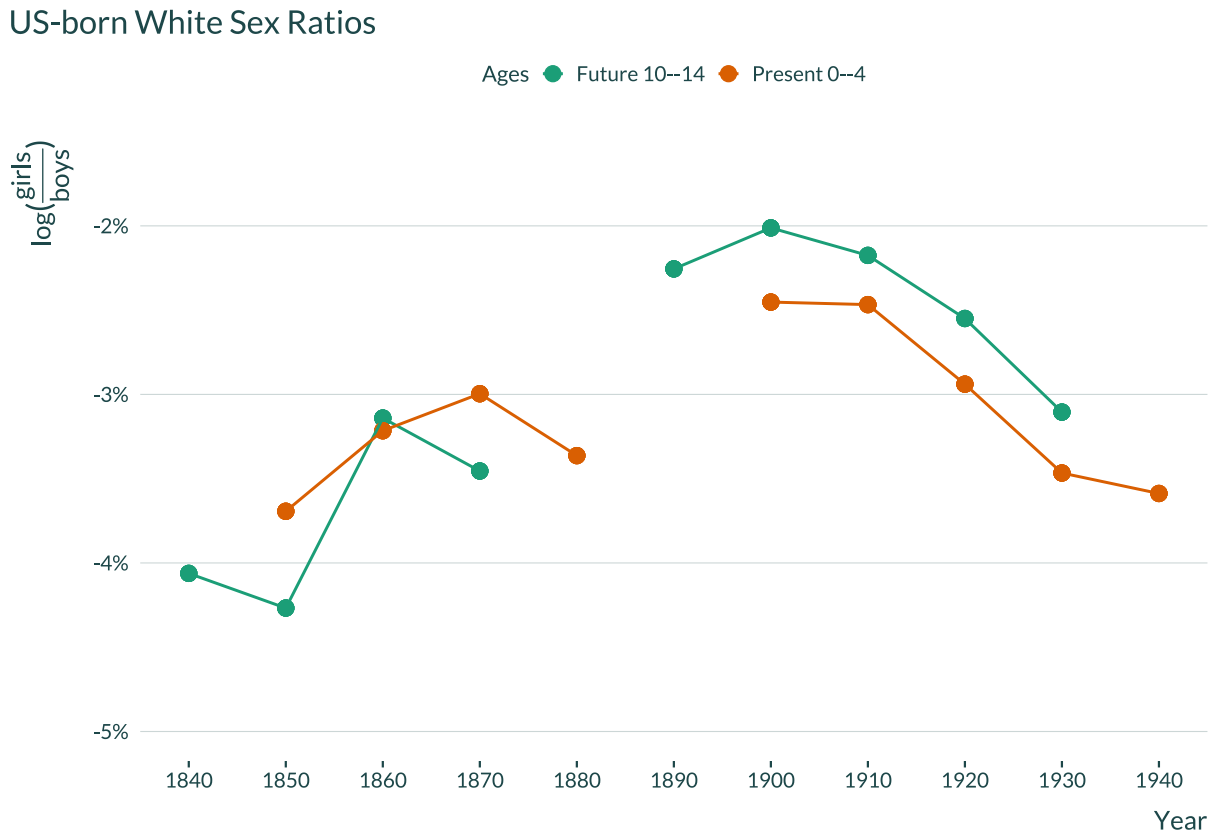


Figure 7: Sex ratios by 5-year cohorts, US native-born whites 1850–1940. The orange line connects the under-5 sex ratio in each census year. The green line connects the sex ratio of 10–14 year olds in the following census year (10 years later). Data from PUMS.

# 7 Discussion

Based on childhood sex ratios, we find that the infant mortality of US whites in the mid-19th century was less than half of the values presented in *HSUS*: 80 rather than 200. While coarse, our estimates nonetheless place US whites among the healthiest populations of the

[47]The 1840 census volume gives an under-5 sex ratio of 5.5% more boys than girls, but the 1850 10–14 cohort has only 4% more boys than girls, suggesting the 1840 value is biased towards boys. Nonetheless, both suggest that infant mortality was even lower in 1840 than in the 1850–1880 period.

[48]Fogel (1986), Pope (1992) and Hacker (2010) all find that life expectancy declined from 1800 to 1850. Similarly, Margo and Steckel (1983) and Komlos (1987) find that adult male heights declined over the same period (often referred to as the 'antebellum puzzle').

19th century. Our results point to a dramatic shift from current views of US population health in the mid-19th century. If we extrapolate our revision in infant mortality to life expectancy using the West model life tables, life expectancy at birth increases 20 years: from 36 to 57.[49] Although dramatically departing from the current view in *HSUS*, our results are not without precedent. They are, for example, consistent with some estimates of life expectancy in the early and mid-19th century from genealogical data.[50] And more narrowly, Lynch et al. (1985: table 4) find that infant mortality in 1850 on the US frontier (Utah) was about 100 deaths per 1000.

Rather than being a revision, our results fill a gap in existing historical knowledge. As discussed above, direct evidence on births and deaths – whether from vital statistics or maternal recall – is extremely limited for the 19th-century US. This has prevented even indirect estimation of infant mortality in all but a few cases.[51] Most work has therefore focused on mortality at older ages, which can be estimated by genealogical records and/or census data (e.g., Kunze 1979: chapter 4; Pope 1992; Kasakoff and Adams 1995; Ferrie 1996). The few rates of infant mortality given for the 19th-century US as a whole are life-table extrapolations from these indirect estimates of mortality at older ages (Haines 1998; Hacker 2010).[52] In place of infant mortality, adult male heights have served as a proxy for population health (e.g., Fogel et al. 1983; Costa and Steckel 1997). This literature corroborates our results in several dimensions. First, as discussed in section 2, 19th-century US whites were tall, suggesting low infant mortality. Second, US heights declined across the second half of the 19th century; we find an increase in infant mortality across the same period. The HSUS series places 19th-century US IMR in the upper range of contemporary Europe. Our results instead agree with height data, suggesting that US white were among the healthiest populations of the 19th century.

Once we break US whites into urban and rural populations, we see that a large part of

[49] Taken from levels 8.5 (mean of the values from levels 8 and 9) and 16.5 (mean of the values from levels 16 and 17) in Coale and Demeny (1983:45–49). This calculation is, of course, highly speculative, and we only offer it as an illustration. As illustrated in Figure 3, infant mortality and mortality at older ages do not necessarily go together, and our revision to infant mortality does not imply anything for mortality at older ages.

[50] Fogel (2004: table 1.1) places life expectancy at birth in 1800 at 56. Pope (1992: table 9.2) and Ferrie (1996: table A1) place life expectancy at age 20 around 45 in the 1850s, which corresponds to life-expectancy at birth in the mid 50s by the West model life table.

[51] Studies are usually limited to times and places where records of births and deaths are available. This is usually specific geographic areas (e.g., Massachusetts, Utah, specific cities) and/or the late 19th century (e.g. Haines 1977; Lynch, Mineau, and Anderton 1985; Preston and Haines 1991; Cutler and Miller 2005; Alsan and Goldin 2019).

[52] Hacker (2010:59) emphasizes that "estimating infant mortality (...) from life expectancy at age 20 (...) is problematic."

the low infant mortality of US whites can be explained by low urbanization. The US was predominantly rural (4/5) in 1860. Contemporary England, for example, was half urban, and infant mortality was around 150–160 during the 19th century, substantially higher than what we find for the US. Taking English values for infant mortality – 220 in urban areas and 100 in rural areas (Woods et al. 1988: table 2), and applying the urbanization rate of the US in 1850, we would get an IMR around 120. So in a sense, we are proposing that the US whites were about 2/3 as unhealthy as their English contemporaries. We disaggregate the US white population into rural and urban, and use equation *R2* to make predictions of infant mortality. As seen in in Figure 8, we find that infant mortality was around twice as high in urban areas relative to rural. US rural infant mortality was about 70, placing rural US whites as slightly healthier than rural Scottish regions (Lee 1991:table 1), and about 30 points below rural England as a whole (Woods et al. 1988). Similarly, we find urban infant mortality of about 130, only 20–30 points below contemporary London, though substantially lower than in the burgeoning industrial cities of England (William 1994; Davenport 2020). Urbanization can thus explain much of the US advantage relative to contemporary Europe.

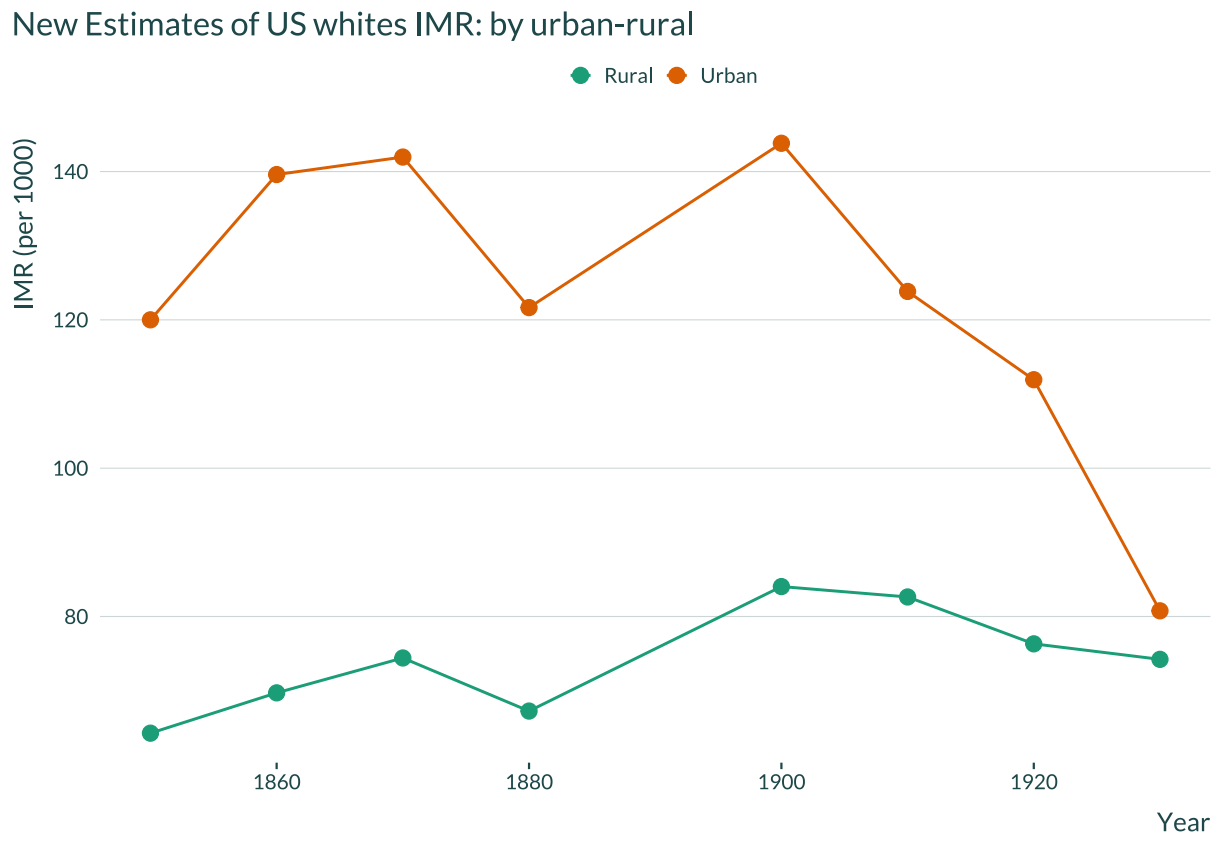


Figure 8: Under-5 sex ratio estimates infant mortality for US whites, urban vs. rural. Calculated via equation *R2*. Sex-ratio data from US census: published volumes and PUMS.

When broken down regionally, our results suggest that 19th-century economic growth was negatively associated with maternal-infant health, likely because of concomitant urbanization. Applying our prediction method (equation *R2*) to regional under-5 sex ratios, we find that from 1850–1880 infant mortality was 30 points greater in the richer, more-industrialized Northeast region than in the poorer, more agrarian Midwest. In Figure 9, we plot regional infant mortality against per-capita income. From 1850–1900, maternal-infant health was negatively associated with per-capita income. This negative relationship can be partly explained by differences in urbanization, which was closely tied to 19th-century economic growth.

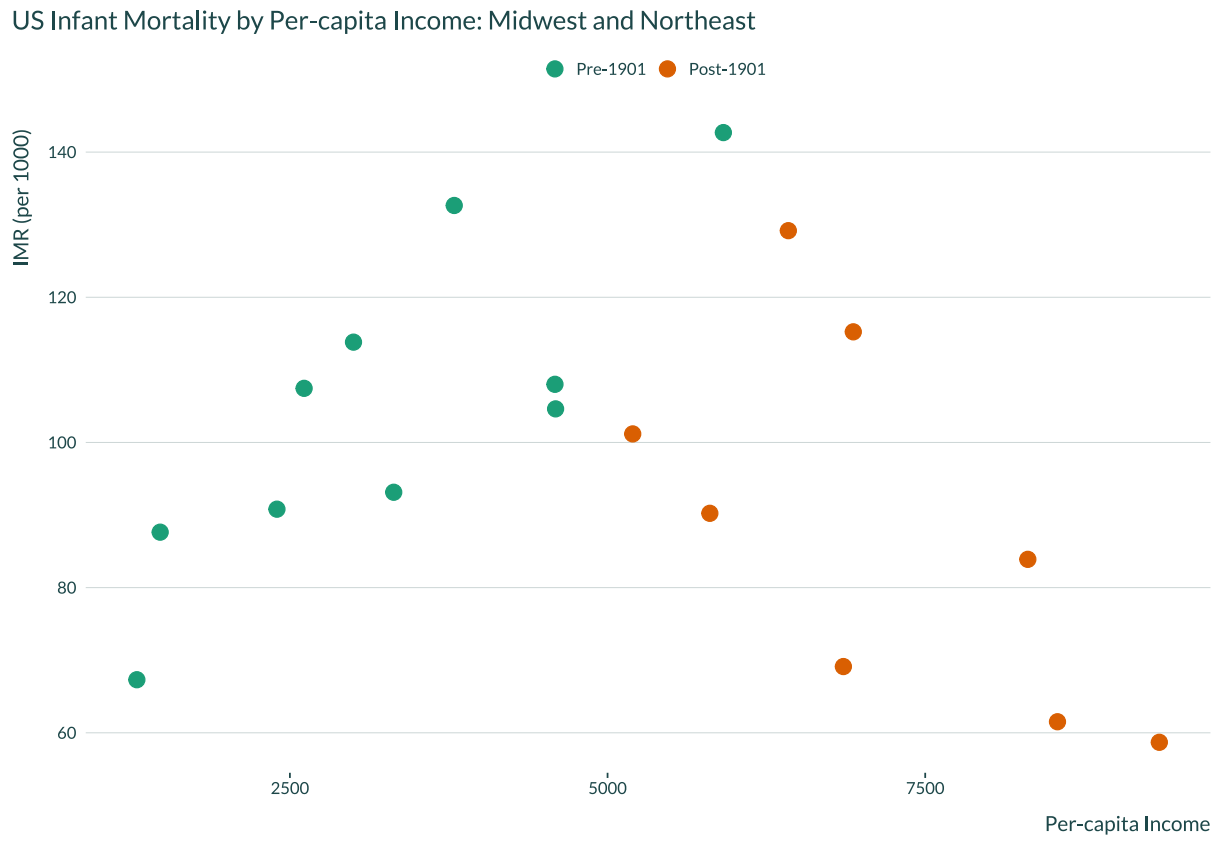


Figure 9: Regional estimates of infant mortality against per-capita income. The dotted lines connect points of each region, the Midwest and the Northeast. IMR estimates calculated from Equation *R2*. Per-capita income estimates from Easterlin (1961:528).

After 1900, the urban penalty began to fall (Figure 7), and the negative association of per-capita income and infant mortality became positive. This reversal, and the decline in urban infant mortality, lined up with the start of a period of investment in public health measures like water filtration and sewage treatment in US cities (Cain and Rotella 2008). Our results thus support the recent consensus that investments in public health were key to

20th century urban infant mortality reductions.[53] Our results in Figure 8 suggest going a step further: until their proceeds were invested in public goods, industrialization and economic development appear to have had a negative impact on US population health,[54] likely because of their association with urbanization.

However, even after accounting for urbanization, US whites still appear to have had lower infant mortality than contemporary Europe. We suggest low inequality as a likely explanation for the good health of US whites. The US, especially the Midwest, was known for having a relatively egalitarian distribution of land ownership.[55] Lining up with this interpretation, the Midwest has the lowest infant mortality. The role of inequality is further supported by Ferrie's (2003:36) finding that infant mortality in rural Illinois was half as much for people with some wealth as opposed to no wealth. We see nutrition as a likely mechanism, as a more egalitarian distribution of land and income would have meant better nutrition for the lower classes. Maternal nutrition is a vital determinant of infant mortality (e.g., Abu-Saad and Fraser 2010), making this a promising channel by which low inequality was translated into good health.[56] This high standard of living in the countryside likely reduced urbanization pressure, leaving better conditions in US cities.

Our proposed narrative is similar to that of Fogel (2004) and reminiscent of elements of Mckeown (1976), in that nutrition takes a leading role in determining historical mortality. However, these authors focused on the role of improved nutrition as caused by economic growth. Our point is that improved nutrition was already possible, but limited by inequality. We can only speculate to what extent similar nutrition, and therefore health, was possible in contemporary Europe under a more egalitarian distribution of land and income. 19th-century variation in infant mortality was huge, larger than the total decline in infant mortality observed in most countries (e.g. Norway 100 vs. Germany 300). How much of this variation could be explained by inequality?

[53] See, e.g., Aykroyd and Kevany (1973); Preston and Van de Walle (1978); Cain and Rotella (2001); Cutler and Miller (2005); Alsan and Goldin (2019).

[54] A point reminiscent of Easterlin (1999).

[55] See, e.g., the discussion of (Engerman, Haber, and Sokoloff 2000). This egalitarian distribution of land, of course, only applied to white settlers, as the land had been violently seized from indigenous communities. See Carlos, Feir, and Redish (2022) for a recent analysis of the close relationship of US property rights and indigenous dispossession. Unfortunately, we lack sufficient census data on the indigenous peoples to include them in our discussion of 19th-century living standards.

[56] 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 IMR (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).

Nowhere is the role of inequality more evident than in the US south, considering racial inequality under slavery. Childhood sex ratios offer a simple, powerful case for US white infant mortality being low in the 19th century. This is certainly not the case for US Blacks. In 1850 and 1860 there were 2% more girls than boys among the US Black population, a value beyond any we observe in European populations.[57] Such female-skewed sex ratios are consistent with the existing HSUS estimate of 340 deaths per 1000 among the Black population in 1850.[58]

Therefore, previous estimates of US infant mortality, by overestimating US white IMR, have profoundly understated the enormous gap in infant mortality between the white and Black populations during slavery. By our new estimates, in the mid-19th century, Black infant mortality was more than 4 times greater than that of whites: 340 vs. 80. The total decline in infant mortality of US whites from 1850 to today – from 80 to 5 – was then less than 1/3 of the raw difference between US whites and Blacks in 1850 – 80 vs 340. Abolition of slavery promised a greater impact on infant mortality than anything offered by the scientific or industrial revolutions, or all the bells and whistles of modern medicine.

# 8 Conclusion

Infant mortality is a key indicator of population health and living conditions more generally. But until now, establishing even approximate levels of infant mortality in the nineteenth-century US has been an intractable problem due to a lack of data. Constructed without vital statistics, or other data on infant deaths or births, existing estimates for the 19th-century US – most notably those published in HSUS (2006) – are conjectural. These estimates place the rate of infant mortality for US whites in the period 1850 to 1880 in the range 175-215 (deaths per thousand live births), with an average of 200. At a time when England's infant mortality was about 150 (similar to that in the State of Massachusetts), the existing estimates are implausibly high relative to known patterns of infant mortality, both within the US and in contemporary Europe.

[57] Childhood sex ratios suggest substantial improvement in Black infant mortality after the abolition of slavery, with the childhood sex ratio moving three points to slightly more boys than girls in 1870. However, deriving infant mortality rates from these sex ratios is not a simple matter. For example, there could be changes in enumeration practices after Abolition. In a separate, ongoing paper, we analyze this sex-ratio data and present new evidence of Black maternal-infant health under slavery and Reconstruction. In this paper, we do not present new estimates of Black infant mortality. It is sufficient for our point to note that the HSUS series is plausible, given evidence from childhood sex ratios.

[58] Much as we have argued that good nutrition contributed to the low infant mortality of 19th-century US whites, see Steckel (1986) for a discussion of maternal malnutrition as a potential cause of the high infant mortality among enslaved Blacks.

We provide 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. Because of the well-known biological survival advantage of infant females, high rates of infant mortality skew the surviving population towards girls (absent extreme sex discrimination). We use historical vital statistics from Europe and the US to characterize the relationship between infant mortality and under-5 sex ratios. Out-of-sample testing and quantile regression enable us to describe the uncertainty in predictions of infant mortality from sex ratios.

Applying our method to 19th-century US census data, we find that US white infant mortality was less than half of previous estimates: about 80 deaths per 1000 in the period 1850–1880, rather than 200. Using hypothesis testing, we reject at the 5% significance level an average infant mortality greater than 125 for US whites across the same period. Our results suggest US whites were among the healthiest populations of the 19th century, with infant mortality substantially below levels in Europe, making it no surprise that millions of voluntary immigrants left Europe for the US during the period. Part of the advantage of the US over Europe can be explained by low rates of urbanization, as the US was predominantly rural in the 19th century. Beyond urbanization, we suggest low inequality as a potential explanation for the relatively low infant mortality of US whites, likely through the channel of maternal nutrition.

Current research on historical infant mortality focuses on explaining the causes behind the 20th century 'mortality revolution.' This research has greatly enriched our understanding of trends in mortality, and our findings fit comfortably into the new consensus which assigns key roles to both nutrition and public health policy in overcoming the mortality challenge posed by urbanization. However, a focus on trends in mortality tends to divert attention from the enormous variation in levels of infant mortality in the 19th-century. The enormous difference in infant mortality across populations makes it an even more important determinant of living standards in the 19th century than we consider it today.[59]

Nowhere were huge differences in infant mortality, and therefore living standards, more evident than with Blacks and Whites within the antebellum US (in the period of slavery). While childhood sex ratios imply a revision of US white infant mortality, they corroborate existing estimates for the Black population. Our finding of relatively low infant mortality among US Whites implies a staggering racial health gap during slavery: infant mortality

[59]For example, the difference between India and the US today – 30 deaths per 1000 vs 5 deaths per 1000 – is minuscule compared to the difference between Norway and Germany in 1850 – 100 deaths per 1000 vs 300 deaths per 1000.

among US whites was less than 1/4 that of US Blacks in the 1850s. Taken as a whole, our results suggest that the 19th-century US was mostly a very healthy place by contemporary standards. However, this good health was restricted along racial lines, painting a jarring picture: in terms of infant mortality, US whites were among the healthiest populations of the 19th-century, while enslaved Blacks were among the most miserable.[60]

[60]Though we focus here on the white and Black population, other non-white groups were also excluded from the good health enjoyed by US whites. Indigenous peoples were violently removed, and the racist immigration policy of the US government literally excluded many non-Europeans from enjoying US living standards (e.g., the Chinese Exclusion Act of 1882).

# 9 References


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. https://doi.org/10.1080/03670244.1973.9990310.

Barber, Rina Foygel, Emmanuel J. Candes, Aaditya Ramdas, and Ryan J. Tibshirani. 2021. "Predictive Inference with the Jackknife+." *The Annals of Statistics* 49 (1): 486–507.

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

Butler, Ronald, and Edward D. Rothman. 1980. "Predictive Intervals Based on Reuse of the Sample." *Journal of the American Statistical Association* 75 (372): 881–89.

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. 2006. "Transmuting the Industrial Revolution into Mortality Decline." In *Demographic Transition Theory*, edited by John C. Caldwell, 157–80. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-4498-4_8.

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. 2006. "Historical Statistics of the United States, Volume One: Population." *New York Cambridge Univ Pr*.

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*. Second. 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.

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. 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.

Costa, Dora, and Richard H. Steckel. 1997. "Long-Term Trends in Health, Welfare, and Economic Growth in the United States." In *Health and Welfare During Industrialization*, 47–90. University of Chicago Press.

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.

Easterlin, R. 1996. "The Nature and Causes of the Mortality Revolution." *Growth Triumphant: The Twenty First Century in Historical Perspective*, 69–82.

Easterlin, Richard A. 1961. "Regional Income Trends, 1840-1950." In *American Economic History*, edited by Seymour E. Harris, 525–47. McGraw-Hill.

———. 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. 2003. "Slavery, Freedom, and Sen." *Feminist Economics* 9 (2-3): 185–211.

Engerman, Stanley L., Stephen H. Haber, and Kenneth L. Sokoloff. 2000. "Inequality, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States." In *In Claude Menard (Ed), Institutions, Contracts and Organizations*, 108–34. Northampton, Mass.: Edward Elgar.

Ferrie, Joseph P. 1996. "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 29 (4): 141–56.

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. https://doi.org/10.2307/203716.

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. 2020. "The Brazilian Bombshell? The Long-Term Impact of the 1918 Influenza Pandemic the South American Way." 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. https://doi.org/http://dx.doi.org.cmich.idm.oclc.org/10.1017/S0145553200010579.

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.

———. 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*, Fourth, 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.

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.

Kasakoff, Alice Bee, and John W. Adams. 1995. "The Effect of Migration on Ages at Vital Events: A Critique of Family Reconstitution in Historical Demography." *European Journal of Population/Revue Européenne de Démographie* 11 (3): 199–242.

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.

———. 1996. "Anomalies in Economic History: Toward a Resolution of the 'Antebellum Puzzle'." *The Journal of Economic History* 56 (1): 202–14.

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. https://doi.org/10.1016/j.ehb.2003.12.006.

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.

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. 2022. "Childhood Sex Ratios Reveal Infant Mortality: Evidence from Historical Populations."

Mitchell, Brian. 1998. *International Historical Statistics: Europe 1750-1993*. Springer.

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

Riley, James C. 2001. *Rising Life Expectancy: A Global History*. Cambridge University Press.

———. 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. https://doi.org/10.1111/j.1728-4457.2005.00083.x.

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.

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.

Steckel, Richard H. 1986. "A Peculiar Population: The Nutrition, Health, and Mortality of American Slaves from Childhood to Maturity." *The Journal of Economic History* 46 (3): 721–41.

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.

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.

Vallin, Jacques. 1989. "La mortalité en Europe de 1720 à 1914 : tendances à long terme et changements de structure par sexe et par âge." *Annales de Démographie Historique* 1989 (1): 31–54. https://doi.org/10.3406/adh.1989.1730.

———. 1991. "Mortality in Europe from 1720 to 1914: Long-term Trends and Changes in Patterns by Age and Sex." *The Decline of Mortality in Europe*, 38–67.

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.

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 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.

Zehetmayer, Matthias. 2011. "The Continuation of the Antebellum Puzzle: Stature in the US, 1847." *European Review of Economic History* 15 (2): 313–27.

# 10 Appendix

## 10.1 US Under-5 Sex Ratios

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

| Year | Rural | Urban | Midwest | Northeast | South | Total | West |
|---|---|---|---|---|---|---|---|
| 1850 | -3.915 | -1.981 | -3.809 | -2.416 | -4.802 | -3.471 | -3.402 |
| 1860 | -3.726 | -1.301 | -3.104 | -2.195 | -4.471 | -3.178 | -3.864 |
| 1870 | -3.563 | -1.219 | -2.994 | -1.542 | -4.782 | -3.119 | -1.888 |
| 1880 | -3.812 | -1.923 | -2.913 | -2.397 | -4.834 | -3.370 | -3.335 |
| 1900 | -3.229 | -1.154 | -2.514 | -1.193 | -3.588 | -2.537 | -2.852 |
| 1910 | -3.278 | -1.848 | -2.634 | -1.662 | -3.573 | -2.674 | -3.034 |
| 1920 | -3.497 | -2.261 | -3.014 | -2.146 | -3.417 | -2.902 | -3.360 |
| 1930 | -3.570 | -3.343 | -3.746 | -3.234 | -3.416 | -3.454 | -3.286 |
| 1940 | -3.506 | -2.961 | -3.652 | -3.893 | -2.331 | -3.253 | -3.141 |
| 1950 | -4.807 | -3.854 | -4.474 | -4.545 | -4.072 | -4.286 | -3.863 |

## 10.2 Robustness

Here we present the IMR predictions of three alternative regressions of infant mortality on childhood sex ratios, as well as the regressions underlying these predictions.

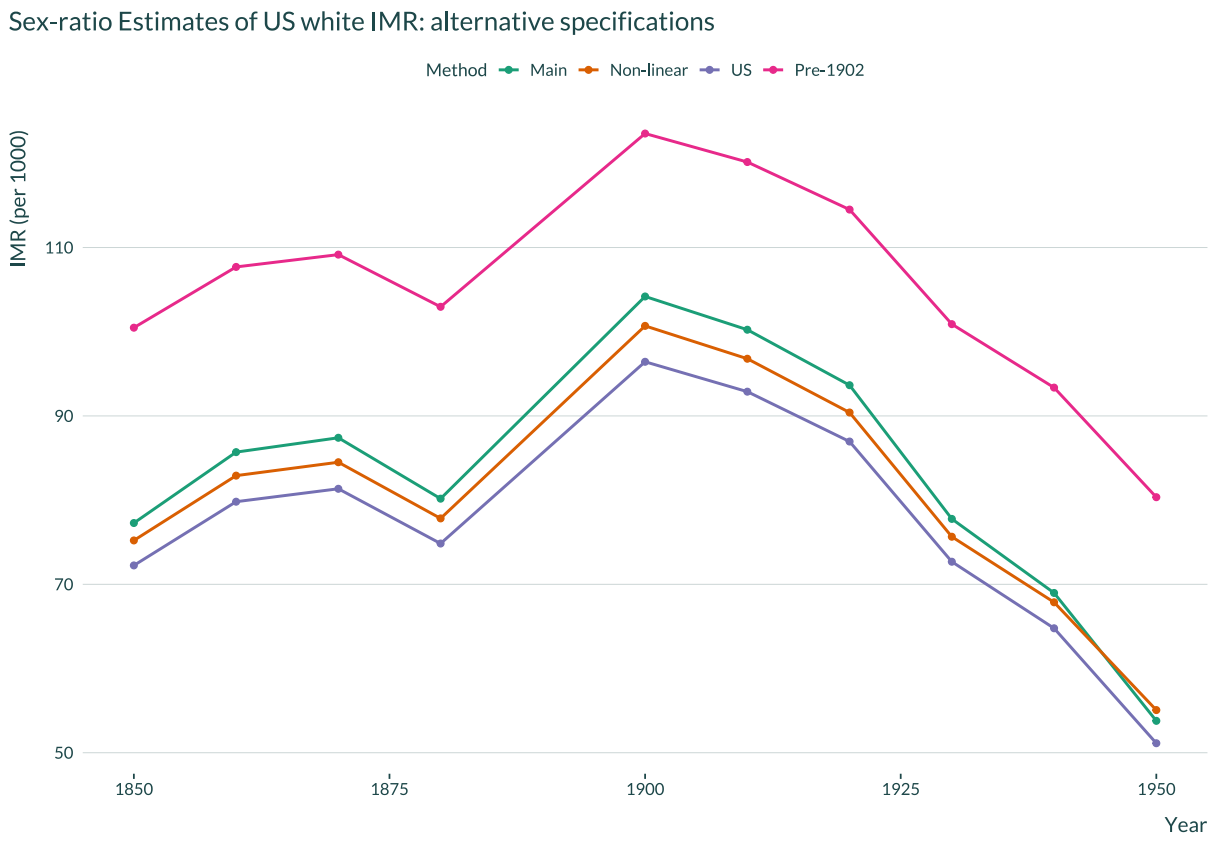


Figure 10: Alternative estimates of US white IMR, based on sex ratios. Main: full-sample, bivariate regression. Non-linear: full-sample, second-order polynomial regression. US: sample restricted to US data only. Pre-1902: sample restricted to observations from 1901 and earlier.

Table 2

| | Non-linear | US | Pre-1902 |
|---|---|---|---|
| (Intercept) | 98.9311 *** | 162.1333 *** | 186.1751 *** |
| | (1.3842) | (5.3691) | (4.0815) |
| poly(sr, 2)1 | 819.3349 *** | | |
| | (26.5415) | | |
| poly(sr, 2)2 | 71.9999 ** | | |
| | (23.0227) | | |
| sr | | 25.9007 *** | 24.6924 *** |
| | | (1.6598) | (1.9274) |
| N | 349 | 139 | 113 |
| R2 | 0.7452 | 0.6918 | 0.5628 |

Standard errors are heteroskedasticity robust. *** p < 0.001; ** p < 0.01; * p < 0.05.

# 11 Software Used

Analysis done in R version 4.2.1 (2022-06-23), with the following packages:

Table 3: R Packages

| Package | Loaded version | Date | Source |
|---|---|---|---|
| dplyr | 1.0.9 | 2022-04-28 | CRAN (R 4.2.1) |
| forcats | 0.5.1 | 2021-01-27 | CRAN (R 4.2.1) |
| ggplot2 | 3.3.6 | 2022-05-03 | CRAN (R 4.2.1) |
| kableExtra | 1.3.4 | 2021-02-20 | CRAN (R 4.2.1) |
| lmtest | 0.9-40 | 2022-03-21 | CRAN (R 4.2.1) |
| mediocrethemes | 0.1.3 | 2022-07-17 | Github (vincentbagilet/mediocrethemes) |
| purrr | 0.3.4 | 2020-04-17 | CRAN (R 4.2.1) |
| readr | 2.1.2 | 2022-01-30 | CRAN (R 4.2.1) |
| sandwich | 3.0-2 | 2022-06-15 | CRAN (R 4.2.1) |
| stringr | 1.4.0 | 2019-02-10 | CRAN (R 4.2.1) |
| tibble | 3.1.8 | 2022-07-22 | CRAN (R 4.2.1) |
| tidyr | 1.2.0 | 2022-02-01 | CRAN (R 4.2.1) |
| tidyverse | 1.3.2 | 2022-07-18 | CRAN (R 4.2.1) |
| zoo | 1.8-10 | 2022-04-15 | CRAN (R 4.2.1) |

# 12 Data Sources

## 12.1 US Data Sources

Data are for the census category "white" unless otherwise specified. Following US vital statistics definitions, urban refers to cities with population 10,000 or more.

### US national data (1850–80, 1900–70)

Under-five white populations by sex for 1900 to 1970 are presented in the *1970 Census of Population: Characteristics of the Population*;[61] values for 1850 to 1870 are in the *Ninth Census – Volume II. The Vital Statistics of the United States;*[62] the published census values for 1880 are from the *1880 Census, Volume 1, Statistics of Population.*[63] IPUMS full count "samples" are available decennially for 1850-1880, and 1900-1940. Where both are available, we average the census volume and IPUMS full count values (taking each as a plausible tally of the underling census manuscripts). As discussed in the text, we do not include 1890 in our analysis because that census's enumeration of ages was inconsistent with that in the other censuses.

Infant mortality rates for the US as a whole (1936–1970) are from U.S. Department of Health, Education, and Welfare (1996). *Vital Statistics of the United States 1992, Volume II – Mortality.*[64]

### The State of Massachusetts (1860–1925)

We use state totals quinquennially 1860-1895, 1905-1915, and 1925. For 1900, 1920, and 1930 we have various regional breakdowns of Massachusetts data (see below), so we do not use the state totals. The state-level data are for the total population (white and nonwhite). Annual infant mortality rates (1856–1925) are from *HSUS (2006)* Series Ab928. Massachusetts state censuses provide under-five populations by sex decennially 1865-1925.[65] The US federal cen-

[61] Chapter B. General Population Characteristics, United States Summary, Table 53 Age by Race and Sex: 1900 to 1970, pp. 276-77; Washington DC: GPO 1973.

[62] Table XXVI – "Ages, with sex at each period of life, of the white population of the United States, at the censuses of 1870, 1860, and 1850," pp. 610, 619.

[63] Table XX, "Population of the United States, by specified age, sex, race ... 1880," p. 548.

[64] Section 2. Infant Mortality, Table 2-2 "Infant, neonatal, and postneonatal mortality rates, by race: Birth-registration States, 1915–32, and United States, 1933-92" (pp. 3–4 of Section 2; pdf pp. 507-08). This is also the source for the HSUS series for this period.

[65] *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*,

suses have data decennially for 1860-1890 and 1910.[66] We average the values from published federal census volumes with the available IPUMS full count data (1860-1880, 1910)[67]

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

Aggregates were formed to achieve a minimum under-5 population over 49,000.

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; 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. 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.

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, infant mortality rates are calculated from on births and infant deaths for 1915–1919, taken from annual reports of birth statistics for the BRA.[68] 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.

---

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

[66] *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.

[67] 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.

[68] 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)

Under-five populations by sex are from the IPUMS 1920 full count data.[69]

US areas in 1930 include 66 observations. 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 (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.[70] 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[71]

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.[72] 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

[69]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.

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

[71]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.

[72]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)

populations by sex are from the IPUMS 1930 full count data.[73]

[73]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.

## 12.2 Non-US Data Sources

The *Human Mortality Database* provides original data and access to other sources for infant mortality rates and under-five sex ratios for many historical populations. 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.[74] In many cases, the data are available from *International Historical Statistics* (Palgrave Macmillan (Ed.) 2013), which we abbreviate as *IHS* below. For infant mortality rates, we rely on official vital statistics except when demographic scholarship offers better estimates. In general, we calculate sex-ratio values (girls/boys in the under-five population) from official population counts by sex and age, most often census counts. For registry-based sex ratios, we take values at five-year intervals. Specific sources and methods by country follow.

### Australia (1876–1971)

Infant mortality rates for 1876–1901 are from McDonald et al. (1987:58).[75] Rates for 1901–1971 are from Australian Bureau of Statistics, Historical Population.[76]

Under-5 populations by sex are census values for non-aboriginal populations. We have decennial data from 1881–1921 and 1961–1971, and single-year values for 1933, 1947, 1954, and 1966.
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.[77]
Data for 1921, 1933, 1947, 1954, 1961, and 1966 are reported in the Census of 1966.[78]
The data for 1971 are calculated from values for the total and aboriginal populations in the 1971 census. Age by sex for the total population is in Part 9 of Bulletin 1, *Summary of Population*.[79] The age-sex data for the Aboriginal population are from Bulletin 9. *The Aboriginal Population*.

[74] The HMD "is limited by design to populations where death registration and census data are virtually complete," but for our analysis we include populations with credible but incomplete infant mortality and sex-ratio data.

[75] Series MFM 154

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

[77] *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.

[78] 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.

[79] *Census of Population and Housing, 30 June 1971, Commonwealth of Australia, Bulletin 1. Summary of Population*, Part 9 Australia, p. 1.

### Austria (1865–1971)

Infant mortality rates (1865–1971) are from *IHS* (2013: 3577,3580,3583), Series A7.

Under-5 populations by sex are for the years 1869 and 1934, and decenially 1880–1910 and 1951–1971. The data for 1869 and 1910 are from *IHS* (2013: 3440), Series A2.[80] The data for 1880, 1890, and 1900 are reported in editions of *Österreichisches statistisches Handbuch*.[81] The data for 1934, 1951, and 1961 are reported in Statistik Austria, *Statistisches Jahrbuch 2010*.[82]

For the years 1865–1910, Austria refers to Austria-Hungary (as in *IHS*); for later years Austria refers to the Republic of Austria (whose area in 1910 had less than 1/4 of the population of Austria-Hungary (*IHS* 2013: 3402, 3440)).

### Belgium (1842–1970)

Infant mortality rates (1842–1970) are HMD data (downloaded on 2021-10-26).

Under-5 populations by sex are census data, decennially 1846–1866 and 1880–1910, with single-years 1930, 1947, 1961, and 1970. The data were obtained through the HMD (downloaded on 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 Beligique*[83] HMD reports that the data for 1930 are in the 1940 volume of *Annuaire Statistique de la Beligique ed du Congo Belge* (pp. 34–35). HMD reports the data for 1947 are published in volume 5 of the 1847 census of Belgium.[84] HMD reports the data for 1961 are published in volume 5 of the 1961 census.[85] HMD reports the data for 1970 are published in volume 5 of the 1970 census.[86]

[80] Austrian provinces of the Hapsburg Empire. The values here are rounded to the nearest thousand; although we prefer unrounded data, we were unable to locate the data in official sources.

[81] For 1880: *1886* p. 3; for 1890: *1893*, p. 6; for 1900: *1909*, p. 7.

[82] 2.08 Bevölkerung 1869 bis 2001 nach fünfjährigen Altersgruppen und Geschlecht (Population 1869 to 2001 by five-year age groups and sex, p. 45

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

[84] 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.

[85] Institut National de Statistique (1965). *Recensement Général de la Population, 31 décembre 1961, tome V, Répartition de la population par âge.* Bruxelles (publisher and pages not given in HMD source notes).

[86] Institut National de Statistique (1974). *Recensement Général de la Population, 31 décembre 1970, tome V, Répartition de la population par âge.* "Population selon l'état civil et par âge". Bruxelles (publisher and pages not given in HMD source notes).

### Denmark (1836–1970)

Infant mortality rates (1836–1970) are HMD data (downloaded on 2021-10-26).

Under-5 populations by sex are quinquennial 1840–1860 and 1910–1970, and decennial 1870–1890. The data were obtained through the HMD (downloaded on 2021-07-01), which identifies the source as Danmarks Statistik.

### England and Wales (1847–1971)

Infant mortality rates (1847–1971) are from Mitchell & Deane (1962:36-37) for 1847-1937 and from *IHS* (2013: 3582, 3587) for 1942-19171.

Under-5 populations by sex for England and Wales are decennial for 1851–1891 and quinquennial for 1901–1971. The decennial data (1851–1891) are from the censuses of England and Wales, as reported in Mitchell & Deane (1962:12). The quinquennial data (1901–1971) are from the Historic Mortality Datasets of the National Archives.[87]

### Finland (1881–1970)

Infant mortality rates (1881–1970) are HMD data (downloaded on 2021-10-26).

Under-5 populations by sex are quinquennial from 1885 to 1970, obtained through the HMD (downloaded on 2022-02-28) and the HMD identifies Statistics Finland as the source of the data.[88]

### France (1897–1968)

Infant mortality rates (1897–1968) are HMD data (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).[89]

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

[88] 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.)

[89] 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

### Germany (1876–1933)

Infant Mortality Rates (1876–1933) are from IHS (2013: 3577, 3580), Series A7.

Under-5 populations by sex are census values, decennially for 1880-1910, with single-years 1925 and 1933. The data are from various years of the *Statistisches Jahrbuch*.[90] IHS (2013:3454, Series A2) also reports these age-sex population data, but rounded to the nearest thousand.[91]

### West Germany (1960–1970)

Infant mortality rates (1956–1970) are HMD data (downloaded on 2021-10-26).

Under-5 populations by sex for 1960, 1965 and 1970 were obtained through the HMD (downloaded on 2021-10-26), which identifies the source as Statistisches Bundesamt.[92]

### East Germany (1960–1970)

Infant mortality rates (1960–1970) are HMD data (downloaded on 2021-10-26).

Under-5 populations by sex for 1964 and 1970 are census data, obtained through the HMD (downloaded on 2021-10-26), which identifies the source as Statistisches Bundesamt[93]

### Italy (1907–1971)

Infant mortality rates (1907–1971) are from Istat (Italian National Institute of Statistics) Time Series.[94]

Under-5 population by sex are decennial 1911–1931 and 1951–1971; also 1936; from Istat, Time Series.[95]

---

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).

[90]The 1880 data are from the 1883 *Statistisches Jahrbuch*, p. 10; 1890 data are from the *1896* volume, p. 5; 1900 from *1903*, p.6; 1910 from *1919,* pp. 6–7; 1925 from *1929*, p. 14; 1933: *1939*, p. 14.

[91]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.

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

[93]The "Data Sources" (https://mortality.org/hmd/DEUTE/DOCS/ref.pdf – login required) on the HMD data page for East Germany gives the source as "Statistisches Bundesamt, ed. (1996). Bevoelkerungsstatistische Uebersichten 1946 bis 1989 (Teil II). Wiesbaden: Arbeitsunterlage. (Sonderreihe mit Beitraegen für das Gebiet der ehemaligen DDR, Heft 28). The reference days were: 1964-12-31, 1971-01-01 and 1981-12-31."

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

[95]Population, Population by age class and sex, aging ratio and dependency ratio at Census from 1861 to

### New Zealand (1863–1971)

Infant mortality rates are for the non-Maori population from 1863–1945 and for the total population from 1947–1970. Data for 1863–1936 are from Stats NZ Store House.[96] The data for 1936–1945 are from The New Zealand Official Year-book 1957.[97] Data for 1947–1971 are for the total population (including Maori), from Stats NZ Inforshare.[98]

Under-5 census populations by sex are for 1867, 1874, and 1881; quinquennially for 1886–1926 and 1951–1971; 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.[99]

Quinquennial data for 1886–1916 are reported in the 1916 census.[100]

The data for 1936, 1945, and quinquennially 1951–1971, are from the Stats NZ Store House.[101]

### Netherlands (1855–1970)

Infant mortality rates (1855–1970) are HMD data (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 1950–1970.

### Norway (1886–1970)

Credible IMR data for Norway start with 1876. 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.[102] Although

2011 according to reference year borders (Table_2.2.1.xls).

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

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

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

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

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

[101] 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").

[102] STATISTISK SENTRALBYRÅ (Oslo 1961): Dette førte til at tallet på registrerte levende fødte og døde barn ble for lavt og tallet på dødfødte for høyt. De gjeldende bestemmelser om hva en skulle forstå med et

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.

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.[103]

Under-five populations by sex are census values, decennially 1890–1930 and 1950–70; and 1946.[104] Data for 1890–1900 are from Statistics Norway (1910).[105] Data for 1910–1930 are reported in the 1930 census.[106] 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*[107]; 1950 from Statistics Norway (1953), *Folketellingen 1950, Hefte 2.*[108]; 1960 from Statistics Norway (1963), *Folketellingen 1960, Hefte 2.*[109]; and 1970 from Statistics Norway (1971)[110](https://www.ssb.no/a/histstat/nos/nos_a448.pdf) (Population by age and marital status 31 December 1970), pp. 24–25.].

### Scotland (1857–1971)

Infant mortality rates (1857–1971) are HMD data (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.[111] Data for 1891–1901 are in the 1901 census.[112] Quinquennial data for 1911 to

levende og dødfødt barn ble imidlertid stadig innskjerpet overfor jordmødrene, og fra 1876 kan en gå ut fra at de tall som står oppført i den offisielle statistikk stort sett gir et riktig uttrykk for forholdet. See also "Preface" (unpaged) regarding Backer's authorship.

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

[104]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).

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

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

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

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

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

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

[111]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.

[112]Scotland Census Office (1903), *Eleventh decennial census of the population of Scotland . . . 1901 . . . Vol*

1936 are from the General Register Office for Scotland.[113] Quinquennial data for 1941 to 1971 are from General Register Office for Scotland.[114]

### 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.[115]

### Sweden (1753–1970)

Infant mortality rates (1753–1970) are from Statistics Sweden.[116]

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.[117] For years before 1860, we use "official" counts reported by Sundbärg (1908:180).[118] 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.[119]

### Switzerland (1875–1970)

Infant mortality rates (1875–1970) are calculated from data on births and infant-deaths from Historical Statistics of Switzerland, Marriage, Birth, and Death.[120] These IMRs are

*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.

[113]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.

[114]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.

[115]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.

[116]Statistical Database, Population, Population statistics, Deaths, Live births, stillbirths and infant mortality rates by sex. Year 1749–2020 (accessed 2022-03-01).

[117]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.

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

[119]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.

[120]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 1950-1970. The data are from Historical Statistics of Switzerland, Population[121]

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

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

**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). Population. In Vamplew, W. (Ed.), *Australians – Historical Statistics* (pp. 32–41). Broadway, New South Wales: Fairfax, Syme & Weldon.

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

McDonald P., Ruzicka, L. & Pyne, P. (1987). Marriage, fertility and mortality. In Vamplew, W. (Ed.), *Australians – Historical Statistics* (pp. 42–61). Broadway, New South Wales: Fairfax, Syme & Weldon.

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.

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.

U.S. 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.

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