Childhood Sex Ratios Reveal Infant Mortality

Evidence from Historical Populations

Jesse McDevitt-Irwin and James R. Irwin*

10 May, 2022

Background: Although a key dimension of population well-being, infant mortality is largely unknown for most of the world before the mid-20th century. Until now, credible estimates of infant mortality have required data on infant deaths or births. We offer a new approach, using childhood sex ratios. We build from an under-appreciated implication of the well-known biological survival advantage of girls: high rates of infant mortality skew the surviving population toward girls.

Objective: To demonstrate that childhood sex ratios can be used to estimate plausible intervals of infant mortality.

Methods: We model the relationship between sex ratios and infant mortality, demonstrating its empirical validity with historical data, primarily from Europe 1870–1970.

Results: Childhood sex ratios and infant mortality are closely related in our historical data, with a correlation of $r = .89$. Using bivariate regression, we predict infant mortality from sex ratios, obtaining an 80% prediction interval of ±37 points (per 1000).

Conclusions: Childhood sex ratios can reveal broad patterns of infant mortality. Except for populations with 'missing women', childhood sex ratio data provide a basis for interval estimates of infant mortality. Widely available from census data, childhood sex ratios can substantially expand our knowledge of maternal-infant health in historical populations lacking vital statistics.

Contribution: Building from well-known demography and biology, we offer a new indirect method of estimating infant mortality, based on childhood sex ratios. Our method offers fairly broad interval estimates of the infant mortality rate, but it requires no data on infant deaths or births.

*Authors' information: Corresponding author is Jesse McDevitt-Irwin, Columbia University (jrm2275@columbia.edu). James R. Irwin is Professor Emeritus, Central Michigan University. 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.

1 Introduction

Infant mortality is a key indicator of population health and living standards, especially historically, when rates of infant mortality varied greatly over time and space.[1] Unfortunately, broad patterns of infant mortality are largely unknown for most of the world before the mid-20th century, due to a lack of vital statistics data.[2] For mortality at older ages, the empirical record since the 19th century is more complete, with indirect estimates based on intercensal survival, using relatively abundant census data.[3] But to date, credible estimates of infant mortality have required data on at least two of infant deaths, births, or infant population.[4] In this paper, we offer a new approach, estimating plausible ranges of infant mortality using census data on childhood sex ratios, without any need for vital statistics.

It has long been known that biologically, girls are less vulnerable than boys to infant mortality.[5] The corollary which we highlight, and build from, is that high rates of infant mortality

[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]Valaoras (1950:253) suggested "Official statistics on infant and childhood mortality are available for about half of the world's population," citing the United Nations, *Demographic Yearbook*, 1948. Coverage falls off further back in time.

[3]See "Estimation of adult mortality using successive census age distributions," chapter IX of UN Population Division (1983).

[4]Data on either infant deaths or births can be combined with counts of infant population to produce credible estimates of the infant mortality rate within a reasonably small range (reflecting the separation factors applied to capture the time-pattern of deaths relative to births). Note that even indirect methods for estimating infant mortality require some data on child births and deaths, for example in the form of summary birth histories. See "Child Mortality" (Chapters 15–18) in Moultrie et al. (2013) for review and explanation of such methods. In the absence of data on infant deaths or births, infant mortality is sometimes extrapolated from indirect estimates of adult mortality using a model life table. For examples, see the work of Haines (1998) and Hacker (2010) on the 19th-century US.

[5]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). Other key contributions include Waldron (1983), Kraemer (2000), and Wells (2000).

tend to skew childhood sex ratios toward females. Absent 'missing girls' (from fatal sex discrimination), the childhood sex ratio in a population reflects, to a large degree, the level of infant mortality.

If a population has very low infant mortality, say 5 (per 1000) like Canada in 2018, the relative numbers of boys and girls during childhood will about the same as the sex ratio at birth.[6] However, with infant mortality rates in the range of 100 to 150 and more, common in late 19th-century Europe, childhood sex ratios will be distinctly more female than the sex ratio at birth. For example, in England in 1901, with infant mortality just over 150, girls outnumbered boys under the age of five. By 1950, infant mortality had plummeted to 20 and England had about 5% more boys than girls.[7]

The biological vulnerability of infant males makes childhood sex ratios a potential signal of infant mortality. However, childhood sex ratios can instead reflect the social vulnerability of girls.[8] Extremes of male-biased sex ratios are familiar (in social science research and popular media) as evidence of 'missing women' (e.g., Coale 1991; Klasen 1994; Klasen and Wink 2003; Sen 1990; Zeng et al. 1993), and extremes of sex discrimination in the allocation of care, nutrition, and resources eliminate the biological survival advantage of girls (e.g., Barcellos et al. 2014; Chen et al. 1981; Das Gupta 1987). In such cases, the biological link between the level of infant mortality and the childhood sex ratio will disappear. Instead, excess female mortality among infants and children skews sex ratios toward boys, regardless of the level of infant mortality. However, because sex discrimination goes against girls, female-skewed childhood sex ratios are still a powerful signal of high infant mortality and maternal distress.

---

[6] In a healthy population, there are typically 5–6% more males born than female (Grech et al. 2002; Maconochie and Roman 1997). Canada in 2018 had 105.1 boys per 100 girls in the population under age 5 (Statistics Canada, Table 17-10-0005-01, Population estimates on July 1st, by age and sex (accessed 2022-04-25); infant mortality rates are given in Statistics Canada, Table 13-10-0712-01).

[7] Office for National Statistics (UK), Vital statistics in the UK: births, deaths and marriages (Dataset), "Deaths" (release date 3 December 2021; accessed 2022-05-03). The National Archives, Historic Mortality: 1901-1992 dataset, RG 69/1, Population, 1901–1992 (NDAD reference: CRDA/20/DS/1/10; downloaded 2020-06-03).

[8] We draw on Thompson's (2021:467) apt formulation: "boys are more biologically vulnerable and girls more socially vulnerable."

Our goal is to demonstrate that childhood sex ratios can reveal broad patterns of infant mortality. We first model the theoretical relationship between childhood sex ratios and infant mortality. Then we characterize their empirical relationship using data from a set of historical populations for which both infant mortality and childhood sex ratios are available. We find a strikingly strong relationship, which enables us to estimate plausible ranges of infant mortality from childhood sex ratios. Widely available from census data for populations lacking vital statistics, childhood sex ratios have the potential to substantially expand historical knowledge of infant mortality.

2 Modeling Childhood Sex Ratios

While largely unexploited to date,[9] the relationship between infant mortality and childhood sex ratios is a standard feature in well-known demographic models. For example, in the widely-used Coale-Demeny West model life tables, 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 shifts 2.14 percentage points away from girls.[10] A stronger effect of infant mortality on childhood sex ratios is found in historical English life tables. Moving from the English life table of 1881–1890 to that of 1950–1952, the infant mortality rate plunges from 146 to 29 and the sex ratio among 5-year-old survivors moves 3.28 percentage points away from girls.[11] The relationship between infant mortality and childhood sex ratios is implicit in existing demographic models, so we next turn to modeling it explicitly.

[9] Excess male infant mortality has been explicitly considered in efforts to estimate numbers of 'missing women', most notably, see Klasen (1994:1064–1065). More recently, arguing for excess female mortality in infancy and childhood in modern Greece, Beltrán Tapia and Raftakis (2021:6–7) plot childhood sex ratios against infant mortality in historical European data. Those plots show sex ratios skewing toward girls as infant mortality rises (Figures 2 & 3). Attention to sex-ratio issues in research bearing on the urgent problem of 'missing women' in contemporary societies likely explains why the potential for childhood sex ratios to reveal maternal-infant health in historical populations has been thus far overlooked.

[10] Coale and Demeny (1983:47,52). The log-change of the sex ratio follows from the ${}_5l_0$ values by sex, moving from level 11 to level 22. Klasen (1994:1064) argues that the Coale-Demeny West model understates girls' biological survival advantage, which strengthens our argument here.

[11] English life table for 1881–90 (published in 1895) p. xiv of Supplement to the fifty-fifth annual report of the Registrar-General of Births, Deaths, and Marriages in England. Part 1. English life table for 1950–52: pp. 31,33 of The Registrar-General's decennial supplement, England & Wales, 1951, Life Tables.

We use life table concepts to delineate the relationship between childhood sex ratios and infant mortality. Abstracting from the details of observed populations, we model the sex ratio at the exact age of 1 year, denoted $SR1$.[12] We define this age-one sex ratio as $SR1 \equiv ln(\frac{l_1^f}{l_1^m})$, where superscripts refer to sex (f or m) and subscripts to age (0 or 1), adopting the notation of Preston et al. (2001:39–42). Let q^f and q^m denote the infant mortality rates of males and females respectively. Defining the (log) sex ratio at birth as $SRB \equiv ln(\frac{l_0^f}{l_m^0})$, a few steps of algebra take us to the following useful representation:[13]

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

Here, the childhood sex ratio is decomposed into two parts: the sex ratio at birth, and a term capturing the differential survival rates of females and males (arising from the sex differential in infant mortality). Thus the childhood sex ratio reflects the effect of differential infant mortality on the sex ratio at birth. It bears emphasis that this effect goes to zero as infant mortality decreases. With the infant mortality rates typical of healthy populations today, childhood sex ratios simply reflect the sex ratio at birth.[14]

In order to clarify the role of the level of infant mortality in determining the childhood sex ratios, we rely on Taylor series approximations. With $ln(1 + q) \approx q$ for small q, we obtain:[15]

[12] We model the sex ratio among the survivors of infancy from a hypothetical large cohort of births. We are working with l_x (survivors at an exact age) and not upper-case ${}_nL_x$ (person-years lived).

[13]

$$SR1 = ln(\frac{l_1^f}{l_1^m}) = ln(l_0^f \cdot (1 - q^f)) - ln(l_0^m \cdot (1 - q^m)) = ln(\frac{l_0^f}{l_0^m}) + ln(1 - q^f) - ln(1 - q^m)$$

[14] For example, suppose female infant mortality is 4 (per 1000) and male infant mortality is twice as much: the survival differential term would be just 0.4% ($ln(\frac{0.996}{0.992})$) and difficult to distinguish from random variation. But with female infant mortality at 120 per thousand and the male rate at 150, the survival differential term would be over 3.4% ($ln(\frac{0.88}{0.85})$).

[15] The first-order Taylor series approximation is biased towards zero, so the equation below model understates the role of infant mortality in determining childhood sex ratios. For the purposes of exposition, the bias is negligible, but may be noticeable in cases of high infant mortality. For example, if $q^m = .22$ and $q^f = .18$, the term $ln(1 - q^f) - ln(1 - q^m)$ is 0.05 but our approximation is 0.04.

$$SR1 \approx SRB + (q^m - q^f).$$

Next define q as overall infant mortality and $\mu \equiv \frac{(q^m - q^f)}{q}$ as excess male mortality.[16] Substituting these terms, we obtain equation (1), which is our theoretical model of the effect of infant mortality on childhood sex ratios:

$$(1) \quad SR1 \approx SRB + \mu \cdot q$$

Equation (1) offers a simple and useful view of how excess male mortality and the infant morality rate interact to skew the childhood sex ratio away from the sex ratio at birth. Intuitively, the greater is excess male mortality, the more that infant mortality skews the sex ratio among survivors. Importantly, this effect is proportional to the level of infant mortality,[17] meaning the effect will be small for populations with low infant mortality.[18] However, at the high rates of infant mortality seen historically around the world, the role of infant mortality in childhood sex ratios promises to be evident.

Equation (1) has appealing simplicity. However, it masks the fact that an underlying factor, maternal health, affects both the sex ratio at birth and infant mortality. A growing body of research demonstrates that insults to maternal well-being push the sex ratio at birth toward females (Almond and Edlund 2007; Catalano 2003; Fukuda et al. 1998).[19] Maternal health and infant mortality are closely connected (e.g., Kramer 1987), so populations with high

[16]Note that μ is well-approximated by the logarithm of the ratio of male to female infant mortality: $\mu \approx ln(\frac{q^m}{q^f})$.

[17]In fact the effect is slightly more than proportional to q, because of our Taylor series approximations (see note above).

[18]For example, consider a population with infant mortality of 5 (per 1000) and another with 15 (per 1000). Even if we assume a high level of excess male mortality, like $\mu = 0.3$, the implied difference in childhood sex ratios – $\mu \cdot q = 0.003$ – is too small to be distinguished from random variation in sex ratios at birth.

[19]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 et al. 1998; Grech and Masukume 2016; Sanders and Stoecker 2015; Shifotoka and Fogarty 2013; Valente 2015). Knowledge of the role of fetal loss in determining the sex ratio at birth goes back at least to Crew (1948:105–108).

infant mortality will tend to have female-skewed sex ratios at birth.[20]

In sum, unhealthy conditions for infants and their mothers push childhood sex ratios towards females through two channels: (1) via the sex ratio at birth and (2) via excess male infant mortality. The practical importance of the two channels for our work is that different populations could have the same childhood sex ratio with different rates of infant mortality. If variation in sex ratios at birth is large enough, it could obscure the effect of infant mortality. With our goal of estimating plausible ranges of infant mortality from childhood sex ratios, we arrive at an empirical question: to what extent do childhood sex ratios reflect infant mortality? In order to address this question, we turn to characterizing the empirical relationship between infant mortality and childhood sex ratios using data from historical populations.

3 Data

In order to demonstrate that infant mortality can be inferred from childhood sex ratios, we assemble credible historical data from vital statistics and censuses. Our sample covers most of Europe, North America, and the non-indigenous populations of Australia, New Zealand and South Africa, roughly spanning 1870–1970.[21] We end our series in the early 1970s; 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 was not yet a factor in sex-ratio patterns.[22]

[20] This point is forcefully developed by Klasen (1994:1064–1066), in the context of estimating the number of "missing women."

[21] See section 9 regarding our sample. In brief, our data cover: Sweden (1757–1970), Denmark (1840–1970), Belgium (1846–1970), England and Wales (1851–1971), the Netherlands (1859–1970), Scotland (1861–1970), New Zealand (1867–1971), Austria (1869–1971), Australia (1880–1971), Germany (1880–1971), Switzerland (1880–1970), Finland (1885–1970), Norway (1890–1970), France (1901–1968), Italy (1911–1971), and South Africa (1918–1921). For the USA, we have Massachusetts (1865–1915), a subset of states (1920–1930), and the USA as a whole (1940–1970).

[22] With the spread of ultrasound since the 1970s (Campbell 2013), sex-selective abortion has emerged as another source of variation in childhood sex ratios, and an important factor for the issue of "missing women".

For each country, we have annual infant mortality, and population by age and sex at various intervals. For a given year, we calculate the under-5 sex ratio, $SR5 = log(\frac{females}{males})$, for children age 0–4. We pair this sex ratio with the 5-year rolling mean of infant mortality, denoted IMR. 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;[23] and pooling across ages reduces the impact of sex-biased age heaping.

4 Results[24]

Figure 1 plots infant mortality against under-5 sex ratios.[25] The empirical relationship is striking: looking from left to right, as the sex ratio moves toward girls, infant mortality climbs.[26] Infant mortality and the under-5 sex ratio are highly correlated ($r = .89$) within our sample, and populations with low infant mortality tend to have some 4–6% more boys than girls, values close to the sex ratio at birth in a healthy population.

We first fit the regression implied by equation (1), regressing the under-5 sex ratio on infant mortality.[27] We obtain:[28]

$$S\hat{R}5 = -.0533 + 0.255 \cdot IMR.$$

Our results suggest that equation (1) provides a reasonable guide to the empirical relationship

[23] 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.

[24] Data work and analysis was done in *R*. Citations for packages can be found in section 10.

[25] Although the primary direction of causality is from infant mortality to sex ratios, we plot the childhood sex ratio on the x-axis because our goal is to predict infant mortality from sex-ratio data.

[26] This correspondence is even closer when considering within-country relationships. The bivariate regression of infant mortality on childhood sex ratios yields $R^2 = .79$. If we allow slopes to vary by country, we obtain $R^2 = .90$. We stick to the most parsimonious specification in this paper, a bivariate regression, in order to demonstrate the broad point that childhood sex ratios can reveal infant mortality. Future applications of our approach to estimating infant mortality rates should exploit within-country relationships when possible.

[27] We use ordinary-least-squares (OLS) regression throughout the paper. Using quantile regression produces very similar results; see section 8.1.

[28] The heteroskedasticity-robust standard errors are 0.000792 for the intercept and 0.00756 for the slope. See section 8.1 for more detailed regression results.

between infant mortality and childhood sex ratios. The regression intercept is close to a healthy sex ratio at birth, with over 5% more boys than girls (Grech et al. 2002; Maconochie and Roman 1997). The slope co-efficient (25.5%) falls within the typical range excess male mortality (Alkema et al. 2014; Hill and Upchurch 1995).

In order to predict infant mortality rates from childhood sex ratios, we reverse equation (1) and take the under-5 sex ratio as the explanatory variable.[29] We obtain:[30]

$$\hat{IMR} = 0.186 + 3.11 \cdot SR5.$$

We plot this line in Figure 1. The regression predicts infant mortality to be 62 points (per 1000) higher when the under-5 sex ratio is 2 percentage points more female, and that a population with equal numbers of boys and girls would have infant mortality of 186.[31]

Our goal is to infer infant mortality from sex-ratio data. To quantify the uncertainty in such inferences, we use 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.[32] The 80% prediction interval from these errors is just under ± 37 points (per 1000), plotted in Figure 1.[33] Further discussion and plots of our prediction errors can be found in section 8.2, which documents the robustness of our results.

[29] Alternatively, we could infer infant mortality from childhood sex ratios from the regression of sex ratios on infant mortality. The calibration literature (Osborne 1991) refers to this as the "classical approach." Our approach, which takes childhood sex ratios as the right-hand-side variable, is known in this literature as "inverse estimation." We obtain similar results with either approach.

[30] The heteroskedasticity-robust standard errors are 0.00368 for the intercept and 0.0962 for the slope. See section 8.1 for more detailed regression results.

[31] We get similar estimates using the regression of the sex ratio on infant mortality: with a sex ratio 2 percentage points more female, infant mortality is predicted to be about 78 points (per 1000) higher, and a population with equal numbers of boys and girls has a predicted IMR of 209.

[32] This is a slight modification of the *cross validation* approach proposed by Butler and Rothman (1980); more recent literature refers to this as a *jackknife* (Barber et al. 2021).

[33] The precise values are $(-36.9, +36.4)$.

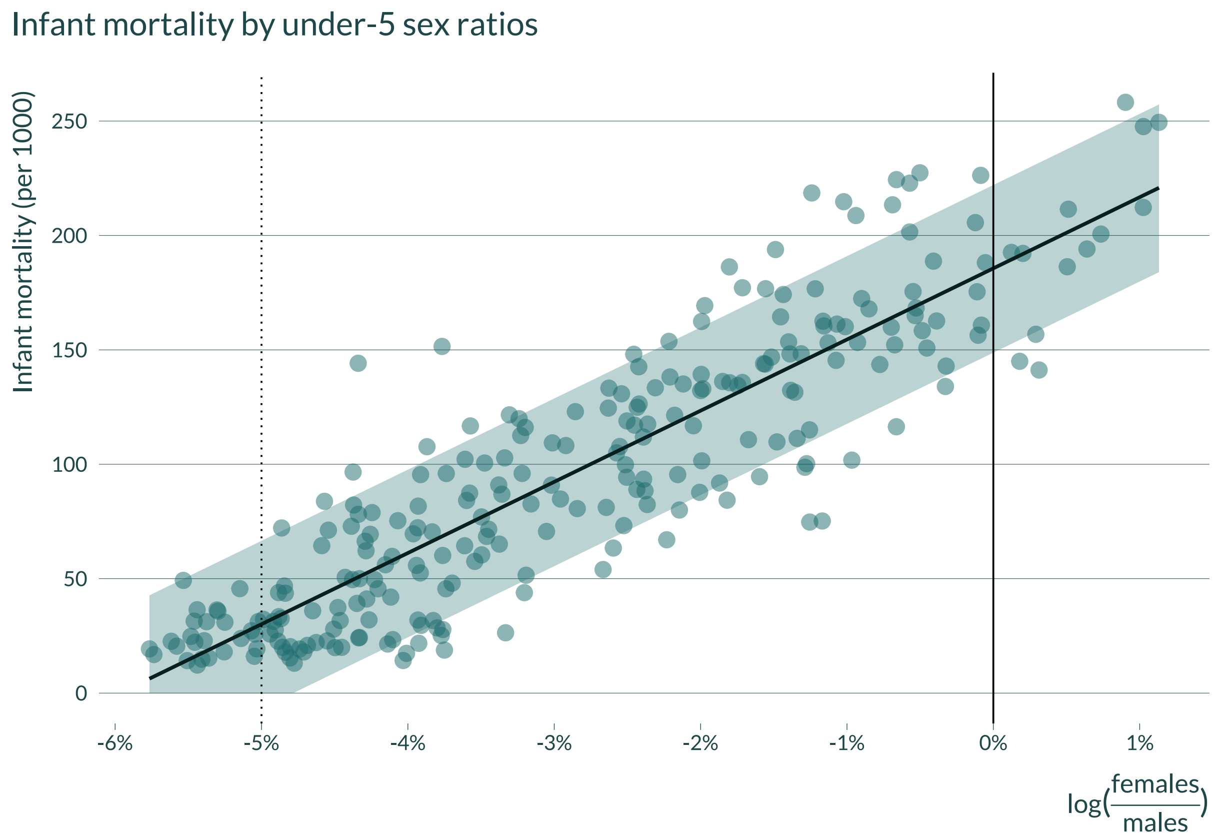

Figure 1: Infant mortality by under-5 sex ratios. The black line is the OLS regression of infant mortality on the under-5 sex ratio. The translucent ribbon is the 80% prediction interval, estimated by out-of-sample testing. The dotted line is a healthy sex ratio at birth of 5% more males than females. Data from Europe and settler colonies.

5 Discussion

Our results show that childhood sex ratios can reveal broad patterns of infant mortality. Our predictions of infant mortality from sex ratios are not precise; for example, even the two-thirds prediction interval is about ±25 points (per 1000). With this wide range of uncertainty, childhood sex ratios will be most informative for populations in which a wide arithmetic range of infant mortality rates are plausible.[34]

The uncertainty in our predictions stems from a number of factors: most simply, random variation and measurement error in both childhood sex ratios and infant mortality. However, we also expect systematic differences across populations, for example due to variation in sex ratios at birth, as discussed in section 2.[35] When the cause of infant mortality is closely related to maternal health, the effect of infant mortality on childhood sex ratios will be reinforced by a female-tilt to the sex ratio at birth. There will be no such effect if the cause of infant mortality is unrelated to maternal health, for example a low incidence of breastfeeding.[36] Thus we can expect different infant mortality rates to be associated with the same childhood sex ratio across populations.

However, our results show that such differences are small relative to the strength of the relationship between infant mortality and childhood sex ratios. With an R^2 of .79, the signal from infant mortality greatly outweighs the noise from other factors, making childhood sex ratios a powerful, unexploited source of information on infant mortality.

[34] For example, we would expect childhood sex ratios to allow us to distinguish between populations with infant mortality of 150 vs. 200 deaths (per 1000), such as mid-19th-century England and the Netherlands. We would not expect sex-ratio data to allow us to distinguish between populations with infant rates of 5 vs. 30, such as Canada and Peru circa 2000 (UN Inter-agency Group for Child Mortality Estimation).

[35] Another contributing factor could be differences in the definition and measurement of 'live births.' Childhood sex ratios are invariant with respect to such registration practices, because they only depend on the surviving population. However, infant mortality rates are sensitive to the extent to which peri-natal mortality is recorded as infant mortality as opposed to stillbirths (e.g., Vallin and Caselli 2006:127). Therefore, differences in measurement practices could produce specious differences in measured infant mortality for a given childhood sex ratio.

[36] Breastfeeding can have huge impacts on infant mortality (Sankar et al. 2015; Wray 1978), but need not be related to maternal health. See for example, Knodel and Van de Walle (1967) on breastfeeding and infant mortality in 19th-century Germany.

As an illustration of the power of sex ratios to reveal infant mortality, consider South Africa in the early-20th century. Vital statistics data indicate an infant mortality rate of some 90 deaths (per 1000) for the "white" population circa 1910 (Mitchell 1998:80). No such data are available for the non-white population of South Africa, but the 1911 census reported population by age and sex and race. For children under the age of 5, there were 6.5% more girls than boys among the non-white population, and 3.1% more boys than girls among whites (Union of South Africa 1912:130–131). This 9.6 percentage-point gap between sex-ratio values suggests a staggering gulf between the two populations in terms of maternal-infant health. As much as 2 percentage points of the gap could be attributed to a difference in the sex ratio at birth between the two groups.[37] The remaining 7.6 percentage-point gap suggests an infant mortality rate of nearly 330 deaths (per 1000) among the non-white population.[38] These values are outside the domain and range of our sample, so the precise result should be treated with caution, but such a female-skewed child population is a powerful signal of high infant mortality and maternal distress.[39]

The South African example illustrates the value of our approach – sex-ratio evidence can offer new insights on infant mortality and maternal health. However, childhood sex ratios will provide misleading estimates of infant mortality in populations where sex discrimination is extreme enough to offset the biological survival advantage of infant girls. Such populations could have healthy-looking childhood sex ratios despite high infant mortality. This possibility – a false positive in a test for low infant mortality – must be considered in applications of

[37]The sex ratio at birth in sub-Saharan African populations tends to be about 2 percentage points more female than in most other populations Klasen (1994:1062). Morse and Luke (2021) offer a compelling argument that this difference reflects maternal health and fetal loss.

[38]We use the slope from our regression to extrapolate from the white infant mortality to an estimated non-white infant mortality. Our finding of a large racial gap in infant mortality is consistent with Mpeta, Fourie, and Inwood's (2018: figure 6) evidence of much smaller stature among black than white men in South Africa in the early twentieth century; both suggest early origins for South Africa's extreme racial inequality. Our finding of very high infant mortality for non-whites contradicts Nattrass and Seekings' (2011:521) conjecture that in the period 1910–1932, "deep poverty was probably generally limited to episodes of drought or disease".

[39]Marco-Gracia and Fourie (2021) argue that the female-skewed childhood sex ratios of non-white South Africans reflect discrimination against young boys. We view high infant mortality and poor maternal health as more plausible explanations.

our approach.

Consider, for example, the Punjab in 1911, where the Census of India reported 6% more boys than girls under the age of 5 (Kaul 1912b:46) – a value which suggests a healthy population. However, both contemporary and modern scholarship make it clear that this preponderance of boys was due to sex discrimination against girls and not to low infant mortality. Visaria (1969:3) notes that the Punjab sex ratios have, historically, been among the most male-biased in India. Kaul (1912a:230–231) attributed this "disparity of the sexes" to the "neglect of female infants," a finding mirrored in the modern literature on sex-biased allocation of household resources (Das Gupta 1987). Indeed, sex discrimination in the Punjab is evident from the 1911 census data, as older sex ratios are increasingly male-biased. With 6% more boys than girls under age 5, there were 16% more boys in ages 5 to 10, and 34% more in ages 10 to 15 (Kaul 1912b:46).

6 Conclusion

Childhood sex ratios can reveal broad patterns of infant mortality. Because girls are hardier than boys, high rates of infant mortality are reflected in female-skewed childhood sex ratios, unless the biological advantage of females is offset by sex discrimination. Assembling historical data from vital statistics, we have shown that plausible ranges of infant mortality can be inferred from sex-ratio data. Often available when vital statistics are not, childhood sex ratios can offer new insights on maternal and infant well-being in historical populations.[40]

[40] Our ongoing work involves using childhood sex ratios to reveal broad patterns of infant mortality and maternal health in the USA and Canada before 1920, 19th-century France, and 19th-century Ireland and Great Britain.

7 References

Alkema, L., Chao, F., You, D., Pedersen, J., & Sawyer, C. C. (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, D., & Edlund, L. (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), 2491–2496.

Barber, R. F., Candes, E. J., Ramdas, A., & Tibshirani, R. J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics*, *49*(1), 486–507.

Barcellos, S. H., Carvalho, L. S., & Lleras-Muney, A. (2014). Child gender and parental investments in India: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, *6*(1), 157–89.

Beltrán Tapia, F. J., & Raftakis, M. (2021). Sex ratios and gender discrimination in Modern Greece. *Population studies*, 1–18.

Butler, R., & Rothman, E. D. (1980). Predictive intervals based on reuse of the sample. *Journal of the American Statistical Association*, *75*(372), 881–889.

Campbell, S. (2013). A Short History of Sonography in Obstetrics and Gynaecology. *Facts, Views & Vision in ObGyn*, *5*(3), 213–229.

Catalano, R. A. (2003). Sex ratios in the two Germanies: A test of the economic stress hypothesis. *Human Reproduction*, *18*(9), 1972–1975.

Chen, L. C., Huq, E., & d'Souza, S. (1981). Sex bias in the family allocation of food and health care in rural Bangladesh. *Population and development review*, 55–70.

Clarke, J. (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, A. J. (1991). Excess female mortality and the balance of the sexes in the population: An estimate of the number of "Missing Females". *The Population and Development Review*, 517–523.

Coale, A. J., & Demeny, P. (1983). *Regional Model Life Tables and Stable Populations* (Second.). Academic Press.

Crew, F. A. E. (1948). *Measurements of the Public Health: Essays on Social Medicine*. Edinburgh: Olvier and Boyd.

Das Gupta, M. (1987). Selective discrimination against female children in rural Punjab, India. *Population and development review*, 77–100.

Davis, D. L., Gottlieb, M. B., & Stampnitzky, J. R. (1998). Reduced ratio of male to female births in several industrial countries: A sentinel health indicator? *Jama*, *279*(13), 1018–1023.

Fukuda, M., Fukuda, K., Shimizu, T., & Møller, H. (1998). Decline in sex ratio at birth after Kobe earthquake. *Human Reproduction (Oxford, England)*, *13*(8), 2321–2322.

Grech, V., & Masukume, G. (2016). The sex ratio at birth in South Africa may be a sentinel health indicator. *Early Human Development*, *103*, 225–227.

Grech, V., Savona-Ventura, C., & Vassallo-Agius, P. (2002). Unexplained differences in sex ratios at birth in Europe and North America. *BMJ: British Medical Journal*, *324*(7344), 1010–1011.

Hacker, J. D. (2010). Decennial life tables for the white population of the United States, 17901900. *Historical methods*, *43*(2), 45–79.

Haines, M. R. (1998). Estimated life tables for the United States, 1850–1910. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *31*(4), 149–169.

Hill, K., & Upchurch, D. M. (1995). Gender differences in child health: Evidence from the demographic and health surveys. *Population and development review*, 127–151.

IGME UN Inter-agency Group for Child Mortality Estimation. (2022). CME Info - Child Mortality Estimates. https://childmortality.org/analysis.

James, W. H., & Grech, V. (2017). A review of the established and suspected causes of variations in human sex ratio at birth. *Early human development*, *109*, 50–56.

Kaul, P. H. (1912a). Census of India 1911. Vol. 14, Punjab. Pt. 1, Report.: Census Reports - 1911.

Kaul, P. H. (1912b). Census of India 1911. Vol. 14, Punjab. Pt. 2, Tables.: Census Reports - 1911.

Klasen, S. (1994). 'Missing women' reconsidered. *World Development*, *22*(7), 1061–1071.

Klasen, S., & Wink, C. (2003). 'Missing women': Revisiting the debate. *Feminist Economics*, *9*(2-3), 263–299.

Knodel, J., & Van de Walle, E. (1967). Breast feeding, fertility and infant mortality: An analysis of some early German data. *Population studies*, *21*(2), 109–131.

Kraemer, S. (2000). The fragile male. *BMJ : British Medical Journal*, *321*(7276), 1609–1612.

Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bulletin of the World Health Organization*, *65*(5), 663–737.

Maconochie, N., & Roman, E. (1997). Sex ratios: Are there natural variations within the human population? *BJOG: An International Journal of Obstetrics & Gynaecology*, *104*(9), 1050–1053.

Marco-Gracia, F. J., & Fourie, J. (2021). The missing boys: Understanding the unbalanced sex ratio in South Africa, 1894–2011. *Economic History of Developing Regions*, 1–19.

Mitchell, B. R. (1998). *International Historical Statistics: Africa, Asia & Oceania, 1750–1993* (3rd ed.). London & New York: Macmillan Reference Ltd.

Morse, A., & Luke, N. (2021). Foetal loss and feminine sex ratios at birth in sub-Saharan Africa. *Population Studies*, 1–16.

Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K., Timæus, I. M., & Zaba, B. (2013). *Tools for Demographic Estimation*. International Union for the Scientific Study of Population.

Mpeta, B., Fourie, J., & Inwood, K. (2018). Black living standards in South Africa before democracy: New evidence from height. *South African Journal of Science*, *114*(1-2), 1–8.

Nattrass, N., & Seekings, J. (2011). The economy and poverty in the twentieth century. In *The Cambridge History of South Africa: Volume 2, 18851994*. Cambridge University Press.

Osborne, C. (1991). Statistical Calibration: A Review. *International Statistical Review / Revue Internationale de Statistique*, *59*(3), 309–336.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishing.

Sanders, N. J., & Stoecker, C. (2015). Where have all the young men gone? Using sex ratios to measure fetal death rates. *Journal of Health Economics*, *41*, 30–45.

Sankar, M. J., Sinha, B., Chowdhury, R., Bhandari, N., Taneja, S., Martines, J., & Bahl, R. (2015). Optimal breastfeeding practices and infant and child mortality: A systematic review and meta-analysis. *Acta paediatrica*, *104*, 3–13.

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

Sen, A. (1990). More than 100 million women are missing. *The New York Review of Books*, *37*(20), 61–66.

Shifotoka, A., & Fogarty, A. W. (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), 327–331.

Struyck, N. (1740). *Inleiding tot de Algemeene Geographie, benevens eenige sterrekundige en andere Verhandelingen*. Amsterdam: Issak Tirion.

Supplement to the fifty-fifth annual report of the Registrar-General of Births, Deaths, and Marriages in England. Part 1. (1895). London: Printed for her majesty's stationery office, by Eyre & Spotswood.

Théré, C., & Rohrbasser, J.-M. (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.

The Registrar-General's decennial supplement, England & Wales, 1951, Life Tables. (1957). London: Her majesty's stationery office.

Thompson, A. L. (2021). Greater male vulnerability to stunting? Evaluating sex differences in growth, pathways and biocultural mechanisms. *Annals of Human Biology*, *48*(6), 466–473.

Union of South Africa. (1912). *Census of the Union of South Africa 1911, Part II, Ages of the People (UG32a-1912).* Pretoria, South Africa: The Government Printing and Stationary Office.

UN Population Division. (1983). *Manual X: Indirect Techniques for Demographic Estimation.* New York: United Nations, Department of Economic and Social Affairs.

Valaoras, V. G. (1950). Refined rates for infant and childhood mortality. *Population Studies*, *4*(3), 253–266.

Valente, C. (2015). Civil conflict, gender-specific fetal loss, and selection: A new test of the Trivers-Willard hypothesis. *Journal of Health Economics*, *39*, 31–50.

Visaria, P. M. (1967). Sex ratio at birth in territories with a relatively complete registration. *Eugenics Quarterly*, *14*(2), 132–142.

Visaria, P. M. (1969). The Sex Ratio of the Population of India, Census of India 1961. *Volume I, Monograph*, *10*.

Waldron, I. (1983). Sex differences in human mortality: The role of genetic factors. *Social Science & Medicine*, *17*(6), 321–333.

Waldron, I. (1998). Sex differences in infant and early childhood mortality: Major causes of death and possible biological causes. In A. D. Lopez & L. T. Ruzicka (Eds.), *Too Young to Die: Genes or Gender?* (pp. 64–83). New York: United Nations.

Wargentin, P. (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.

Wells, J. C. (2000). Natural selection and sex differences in morbidity and mortality in early life. *Journal of theoretical Biology*, *202*(1), 65–76.

Wray, J. D. (1978). Maternal nutrition, breast-feeding and infant survival. In *Nutrition and human reproduction* (pp. 197–229). Springer.

Zeng, Y., Ping, T., Baochang, G., Yi, X., Bohua, L., & Yongpiing, L. (1993). Causes and implications of the recent increase in the reported sex ratio at birth in China. *Population and development review*, 283–302.

8 Appendix

8.1 Regression Results

Here we have the detailed regression results for infant mortality and childhood sex ratios (data from Europe and settler colonies, see section 3).

OLS Regression of IMR on SR5: $\hat{IMR} = 0.1856 + 3.110 \cdot SR5$ ($N = 254$ and $R^2 = .7922$). With robust standard errors of 0.003678 for the intercept and 0.09616 for the slope.[41]

Quantile (median) regression of SR5 on IMR: $\hat{IMR} = 0.1877 + 3.165 \cdot SR5$ ($N = 254$ and $R1 = 0.5785$). With robust standard errors of 0.004089 for the intercept and 0.1157 for the slope.[42]

OLS Regression of SR5 on IMR: $\hat{SR}5 = -0.05326 + 0.2547 \cdot IMR$ ($N = 254$ and $R^2 = .7922$). With robust standard errors of 0.0007920 for the intercept and 0.007563 for the slope.[43]

Implied prediction of IMR from SR5: $IMR = 0.2091 + 3.926 \cdot SR5$

Quantile (median) regression of SR5 on IMR: $\hat{SR}5 = -0.05381 + 0.2599 \cdot IMR$ ($N = 254$ and $R1 = 0.5813$). With robust standard errors of 0.00110 for the intercept and 0.01051 for the slope.[44]

Implied prediction of IMR from SR5: $IMR = 0.2070 + 3.847 \cdot SR5$

[41] Robust standard errors are "HC1" from the *sandwich* package in *R*.
[42] Robust standard errors are Huber ("nid") from the *quantreg* package in *R*.
[43] Robust standard errors are "HC1" from the *sandwich* package in *R*.
[44] Robust standard errors are Huber ("nid") from the *quantreg* package in *R*.

8.2 Prediction Error Plots

Figures 2 and 3 offer two views of our cross-validation prediction errors. As discussed in section 4, 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.

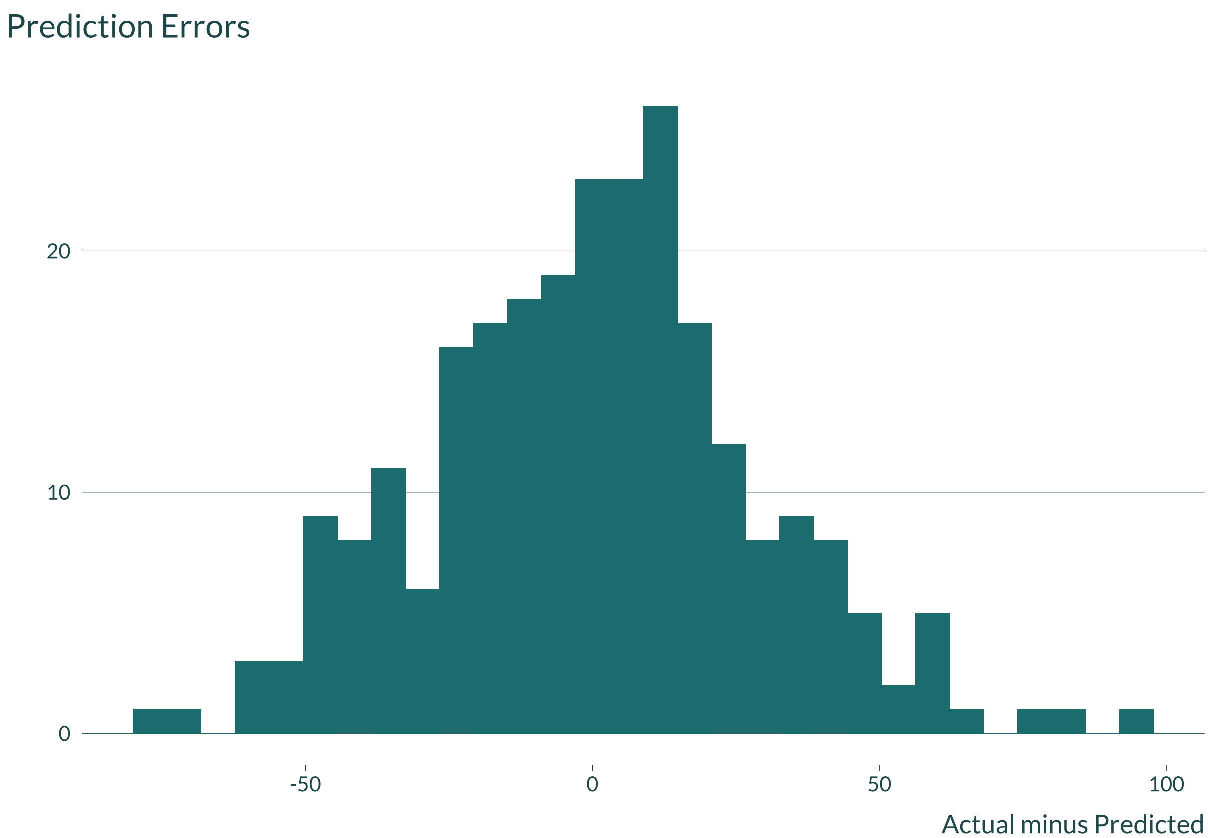

Figure 2: Histogram of absolute prediction errors (Actual minus Predicted) calculated by out-of-sample testing. Data from Europe and settler colonies.

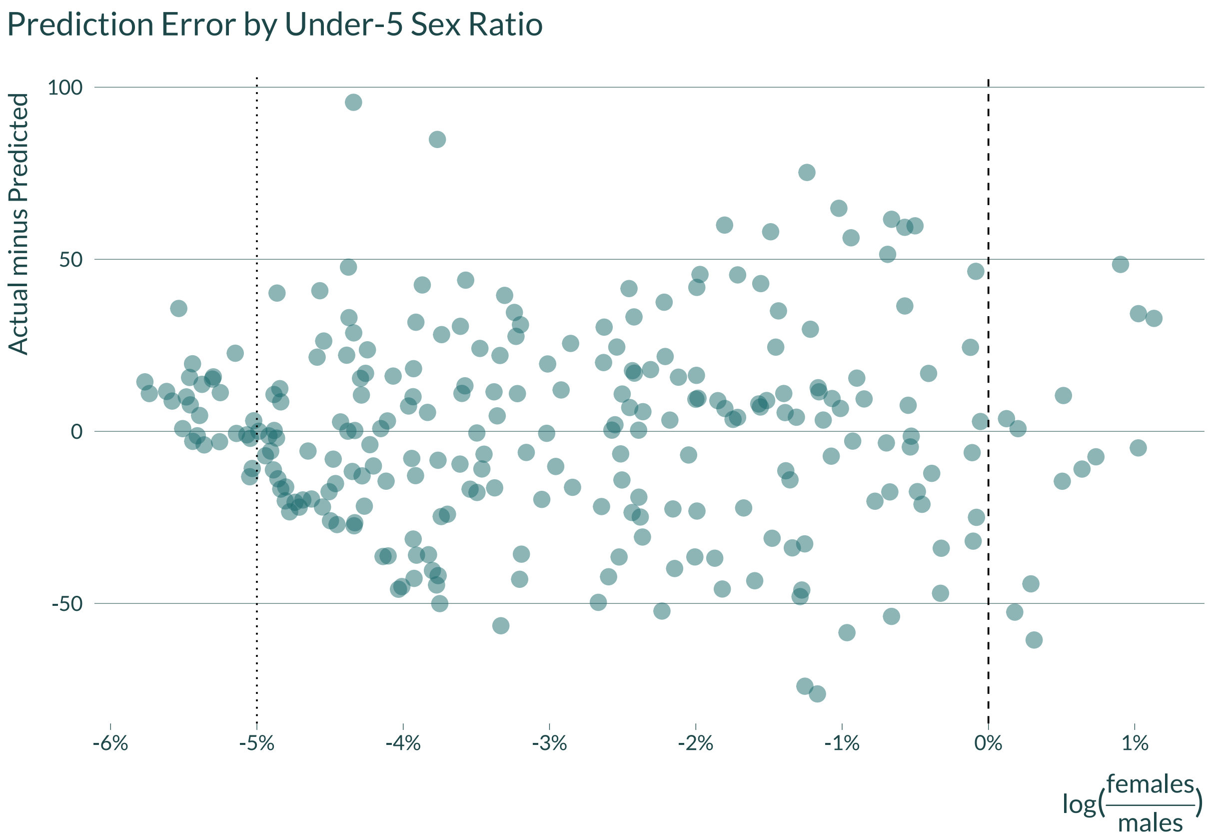

Figure 3: Absolute prediction errors (Actual minus Predicted) by under-5 sex ratio. Prediction errors calculated by out-of-sample testing. Data from Europe and settler colonies.

9 Appendix: 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.[45] 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).[46] Rates for 1901–1971 are from Australian Bureau of Statistics, Historical Population.[47]

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.[48]
Data for 1921, 1933, 1947, 1954, 1961, and 1966 are reported in the Census of 1966.[49]
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*.[50] The age-sex data for the Aboriginal population are from Bulletin 9. *The Aboriginal Population*.

[45]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.

[46]Series MFM 154

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

[48]*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.

[49]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.

[50]*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.[51] The data for 1880, 1890, and 1900 are reported in editions of *Österreichisches statistisches Handbuch*.[52] The data for 1934, 1951, and 1961 are reported in Statistik Austria, *Statistisches Jahrbuch 2010*.[53]

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*[54] 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.[55] HMD reports the data for 1961 are published in volume 5 of the 1961 census.[56] HMD reports the data for 1970 are published in volume 5 of the 1970 census.[57]

[51] 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.

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

[53] 2.08 Bevölkerung 1869 bis 2001 nach fünfjährigen Altersgruppen und Geschlecht (Population 1869 to 2001 by five-year age groups and sex, p. 45

[54] 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)

[55] 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.

[56] 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).

[57] 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.[58]

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.[59]

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).[60]

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

[59] 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.)

[60] 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*.[61] IHS (2013:3454, Series A2) also reports these age-sex population data, but rounded to the nearest thousand.[62]

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.[63]

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[64]

Italy (1907–1971)

Infant mortality rates (1907–1971) are from Istat (Italian National Institute of Statistics) Time Series.[65]

Under-5 population by sex are decennial 1911–1931 and 1951–1971; also 1936; from Istat, Time Series.[66]

---

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).

[61] 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.

[62] 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.

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

[64] 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."

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

[66] 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.[67] The data for 1936–1945 are from The New Zealand Official Year-book 1957.[68] Data for 1947–1971 are for the total population (including Maori), from Stats NZ Inforshare.[69]

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.[70]

Quinquennial data for 1886–1916 are reported in the 1916 census.[71]

The data for 1936, 1945, and quinquennially 1951–1971, are from the Stats NZ Store House.[72]

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.[73] Although

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

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

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

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

[70] *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."

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

[72] 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").

[73] 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

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.[74]

Under-five populations by sex are census values, decennially 1890–1930 and 1950–70; and 1946.[75] Data for 1890–1900 are from Statistics Norway (1910).[76] Data for 1910–1930 are reported in the 1930 census.[77] 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*[78]; 1950 from Statistics Norway (1953), *Folketellingen 1950, Hefte 2.*[79]; 1960 from Statistics Norway (1963), *Folketellingen 1960, Hefte 2.*[80]; and 1970 from Statistics Norway (1971)[81](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.[82] Data for 1891–1901 are in the 1901 census.[83] Quinquennial data for 1911 to 1936

med et 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.

74 Statistisk sentralbyrå, Historisk statistikk, 3.13 Folkemengde, fødte, døde, ekteskap, flyttinger og folketilvekst.

75 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).

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

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

78 *Folkemengden etter kjønn, alder og ekteskapelig stilling, . . .*, Tabeller p. 2.

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

80 *Folkemengden etter kjønn, alder og ekteskapelig status*.

81 *Folkemengden etter alder og ekteskapelig status 31. desember 1970*

82 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.

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

are from the General Register Office for Scotland.[84] Quinquennial data for 1941 to 1971 are from General Register Office for Scotland.[85]

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.[86]

Sweden (1753–1970)

Infant mortality rates (1753–1970) are from Statistics Sweden.[87]

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.[88] For years before 1860, we use "official" counts reported by Sundbärg (1908:180).[89] 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.[90]

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.[91] 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.

[84]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.

[85]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.

[86]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.

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

[88]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.

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

[90]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.

[91]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[92]

United States (1915–1970)

The 1920 and 1930 data are for the states of the Birth Registration Area (BRA) of those years. The US census data for these years refers to populations as of April 15; for an appropriate average IMR to associate with the April 15 U5 populations, we take an average across the 6 years up to the census year, with year t-6 weighted ¾ of one-fifth, year t weighted ¼ of one-fifth, and the other 4 years each weighted one-fifth (thus we treat April 15 as one-quarter through the year). The states included for 1915-1920 were DC ME MA MI MN NH NY PA VT (RI was also in the BRA at this time, but IMR data were missing for 1919 & 1920. The states included for 1925-1930 were the 1915-1920 set plus CA IL IN IA KS MI MT NE NJ ND OH UT WA WI WY DE FL KT MD NC VA WV and RI. In addition to the data sources (below), see U.S. Department of Health, Education, and Welfare (1954) on the "History and organization of the vital statistics system" in the US. Infant mortality data for 1915–1920 and 1925–1930 are from Linder & Grove (1947).[93] We use birth counts to aggregate state-level rates; the births data are from the annually published *Mortality Statistics* (available online from the National Center for Health Statistics). 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.*[94]

Under-5 populations by sex are census data. For 1920 and 1930, the data by state are published in the 1930 censusˆ[US Department of Commerce, Bureau of the Census, *Fifteenth Census of the United States: 1930. Volume 2.* Population, General Report, Statistics by Subjects, Table 24. Age by 5-year periods, by color, nativity, and sex, by divisions and states: 1930 and 1920 (pp. 611–659),

Decennial data for the US from 1940 to 1970 are reported in the 1980 Census.[95]

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

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

[93] *Vital Statistics Rates in the United States 1900-1940,* by F. E. Linder & R. D. Grove (National Office of Vital Statistics, 1947), Table 28, pp. 585-605 (available online from the NBER).

[94]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).

[95]*1980 Census of Population, General Population Characteristics, United States Summary, Table 45. Age by Race and Sex: 1910 to 1980 (p. 1–42).

Massachusetts (1861–1915)

Infant mortality rates (1861–1915) are from *Historical Statistics of the United States* (US Bureau of the Census, 1975:57).[96]

Under-five populations by sex, quinquenially 1865 to 1915, are census data. These are decennial data for 1865–1915 from published volumes of the State of Massachusetts census and decennial data for 1970–1910 from the published volumes of the US Census.

The sources of the age-sex data from the Massachussets state census are as follows: *Abstract of the Census of Massachusetts, 1865*.[97]. *The census of Massachusetts: 1875. Volume I. Population and social statistics*.[98] *The census of Massachusetts: 1885. Volume I. Population and social statistics, Part 1*.[99] *Census of the Commonwealth of Massachusetts: 1895. Volume II. Population and social statistics*.[100] *Census of the Commonwealth of Massachusetts 1905, volume 1, population and social statistics*.[101] *The decennial census 1915*.[102]

The sources of the age-sex data for Massachusetts from the US census are as follows: *Ninth census – volume II. The vital statistics of the United States (June 1, 1870)*.[103] *Statistics of the population of the United States at the tenth census (June 1, 1880)*.[104] *Report on the population of the United States at the eleventh census: 1890, Part II*.[105] *Twelfth census of the United States, taken in the year 1900, Population Part II* (Census Reports Volume II).[106] 1910 Census, Volume 1. Population, general report and analysis.[107]

[96] Chapter B. Vital statistics and health and medical care, Series B148. Infant mortality rate for Massachusetts: 1851–1970. Available online from the US Census Bureau.

[97] Table 1. "Census of Massachusetts, 1865, Distinguishing by Age and Sex, the Number of Inhabitants", p. 2

[98] Ages, p. 269, corrected: the published total for age-one females is 15589 which is an error; the sum of the county values (pp. 263-68) of age-one females is 13589 (the published male total is correct, 13825).

[99] "Ages: under 1 to 80 years and over", p. 434.

[100] "Ages: by five-year periods (by sex)", p. 422

[101] "Ages: under 1 to 80 years and over", p. 480.

[102] Table 25. Ages by native and foreign born and sex, and native and foreign born by color or race, and sex, for the state", p. 478.

[103] The tables of ages. Table XXIII. "Ages with sex for each period of life, of the aggregate population of the United States, by states and territories, 1870–1850", p. 563.

[104] Table XXI. "Population, by Specified Age, Sex, Race, and General Nativity of the Whites, by States and Territories: 1880", p. 592.

[105] Table 3. "Ages by periods of years of the aggregate population, classified by sex, by states and territories: 1890", pp. 104–105.

[106] Ages, Table 3. "Ages by periods of years of the aggregate population, classified by sex, by states and territories: 1900", pp. 110–111.

[107] *Thirteenth census of the United States taken in the year 1910, volume 1, population 1910, General Report and Analysis*, Table 43 "Distribution by age periods of the population, and by each year of age for persons under 25, by divisions and states: 1910", p. 380.

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*.

10 Software Used

Analysis done in R version 4.2.0 (2022-04-22), with the following packages:

Table 1: R Packages

Package	Loaded version	Date	Source
dplyr	1.0.9	2022-04-28	CRAN (R 4.2.0)
forcats	0.5.1	2021-01-27	CRAN (R 4.2.0)
ggplot2	3.3.6	2022-05-03	CRAN (R 4.2.0)
jtools	2.2.0	2022-04-25	CRAN (R 4.2.0)
kableExtra	1.3.4	2021-02-20	CRAN (R 4.2.0)
lmtest	0.9-40	2022-03-21	CRAN (R 4.2.0)
mediocrethemes	0.1.3	2022-05-08	Github (vincentbagilet/mediocrethemes)
purrr	0.3.4	2020-04-17	CRAN (R 4.2.0)
RColorBrewer	1.1-3	2022-04-03	CRAN (R 4.2.0)
readr	2.1.2	2022-01-30	CRAN (R 4.2.0)
sandwich	3.0-1	2021-05-18	CRAN (R 4.2.0)
stringr	1.4.0	2019-02-10	CRAN (R 4.2.0)
tibble	3.1.7	2022-05-03	CRAN (R 4.2.0)
tidyr	1.2.0	2022-02-01	CRAN (R 4.2.0)
tidyverse	1.3.1	2021-04-15	CRAN (R 4.2.0)
zoo	1.8-10	2022-04-15	CRAN (R 4.2.0)