Sex Ratios Reveal Infant Mortality

Evidence from Historical Vital Statistics

Jesse McDevitt-Irwin and James R. Irwin*

31 January, 2022

Abstract

The relative number of girls and boys in a population (i.e. sex ratios) can reveal broad patterns of infant mortality. While familiar as indicators of missing women, sex ratios also reflect infant mortality and maternal health—making them a promising health indicator for populations without vital statistics. We outline the demography and biology of sex ratios, and their relation to maternal health and infant mortality. Using historical data from Europe and its settler colonies, we show that under-five sex ratios can predict infant mortality and quantify the uncertainty in such predictions. Being widely available from census data, sex ratios can shed new light on infant mortality in a wide range of historical and contemporary settings.

A work in progress. 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. We would like to thank Doug Almond, George Alter, Lou Cain, Joel E. Cohen, Kris Inwood, Samuel Preston, and Rodrigo Soares for helpful comments on both this draft and a previous, related manuscript. Jesse McDevitt-Irwin would like to thank his lab-mates Claire Palandri and Vincent Bagilet, as well as Jeffrey Shrader and the participants at his lab for helpful comments.

1 Introduction

Infant mortality—the probability of a child dying before the age of one—is a key indicator of population health and living standards (Reidpath and Allotey 2003; OECD and WHO 2020), as well as a driving force in demographic dynamics (Coale 1989). Yet, the level of infant mortality remains virtually unknown for most of the world before 1950, our understanding severely constrained by a lack of data. Developing and maintaining civil registration and vital statistics remains a continuing challenge today, with more than 30% of births going unregistered (World Bank).

We present a novel set of indicators for infant mortality—child sex ratios.[1] Given population counts by age and sex, widely available in published census reports, we can calculate the relative number of girls and boys under the age of five, and use this sex ratio to infer the level of infant mortality. For populations without vital statistics, where we have little to no direct evidence of infant mortality, sex ratios can shed crucial new light on infant health and maternal well-being. For example, India in the late-19th century, much of Latin America in the early 20th, and North America before the 1920s all had censuses but lacked vital statistics. Sex ratios have the potential to substantially advance our understanding of maternal-infant health during industrialization, colonization and modernization.[2]

Our goal is to demonstrate that child sex ratios reflect infant mortality, and that sex ratios can be used to *predict* infant mortality. We assemble historical data from Europe and settler colonies on infant mortality and sex ratios by age, restricting attention to data which we deem reliable. A simple scatter-plot illustrates the close relationship ($R^2 = 81\%$) between the infant mortality and child sex ratios. Using out-of-sample testing, we predict infant mortality from sex ratios, and quantify the uncertainty in these predictions. We conclude that sex ratios can reveal broad patterns of infant mortality and shed new light on times, places and populations previously understudied. As an illustration, we apply our results to the 1911 census of South Africa. We estimate that infant mortality among black South Africans was on the order of 350-400 deaths per 1000 births (four times that of the white population).

The remainder of this essay goes as follows: Section 2 outlines the demography and biology behind the relationship between infant mortality and sex ratios, Section 3 analyzes this relationship using data from the 19th and 20th centuries, Section 4 discusses these results, and Section 5 underlines the potential usefulness of sex ratios for revealing patterns of maternal health and infant mortality in hitherto under-studied populations.

[1] In the course of our research we have found several papers which pay attention to sex-ratio data as evidence of maternal or infant health. Tavassolli et. al. (2021) seek to explore effects of industrial pollution on "fetal death and infant mortality", highlighting male-biased fetal loss from environmental shocks. The work of Beltrán Tapia and co-authors (2017, 2020, 2021a, 2021b) makes extensive note of the relationship between infant mortality and child sex ratios. However, the authors do not pursue the potential of sex ratios for revealing infant mortality and maternal health. Instead, they focus on detecting sex discrimination against girls, making their work a natural complement to our focus on infant mortality and maternal health.

[2] We use the term "maternal-infant health" to capture the nexus of maternal health, *in-utero* conditions, and infant health. These are intimately connected, and form a "critical phase" of human development (Darling et al. 2020).

2 Background

2.1 Sex ratios and Infant Mortality

Sex ratios are familiar as indicators of missing women,[3] but they are also closely tied to infant mortality. Scholars have long known that boys are both born and buried more often than girls.[4] In a healthy population, there are 5-6% more male live births than female (Maconochie and Roman 1997; Grech, Savona-Ventura, and Vassallo-Agius 2002), but excess male mortality in infancy then drives the sex ratio back towards parity. Excess male mortality is greatest in the first year of life, with mortality typically 15-30% greater than that of females (Drevenstedt et al. 2008; Alkema et al. 2014).[5]

Because of the lower mortality of infant girls, high rates of infant mortality skew the surviving population toward females. 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 be nearly the same as it was at birth. If, on the other hand, a population has an infant mortality rate of 200 per 1000, as was common in 19th-century Europe, then the observed sex ratio during infancy will be distinctly more female than it was at birth, male infants having been culled from the population at a greater rate than females. This effect is central to our analysis, and it is evident in model life tables. For example, in the Coale-Demeny Model West life tables, a workhorse of applied demography, moving from level 22 to level 11, infant mortality increases from 27 to 160 (per 1000), and the under-1 sex ratio moves 1.55 percentage points toward girls.[6]

2.2 Sex Ratios and Maternal Health

Beyond the effect of infant mortality, the sex ratio at birth reflects maternal health. Put simply, poor maternal health tends to tilt the sex-ratio at birth toward females.[7] As said above, in healthy populations, there are about 5% more male live births than female, but maternal stress, pollution and poverty all lead to relatively more female births.[8] Therefore, female-skewed sex ratios are signals not only of infant mortality but also maternal health.[9] Infant health is closely tied to maternal health, meaning the relationship we observe in Section 3 is the result of the combined effect of maternal health and infant mortality on sex ratios.

[3] Among many possible, see e.g. Sen (1990), Klasen (1994), Klasen & Wink (2002) for research on missing women.

[4] This knowledge was nicely summarized in the report on the 1861 census of Scotland: "At birth 106 males appear for every 100 females ... yet, by the law of nature, a law seen in still more powerful operation even in intra uterine life, the tendency to succumb under disease is so much greater in the male than in the female, that by about the 18th year of life the proportion of the sexes becomes equalized, 100 males being alive at that age for every 100 females." These points have been noted since at least John Graunt (1662), followed by Arbuthnot (1710), and Clarke (1786) (cited in Bakwin, 1929). See Grech (2020) for "historical aspects" of the sex-ratio at birth (and excess male mortality, to a lesser extent).

[5] The female mortality advantage in infancy is due 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).

[6] Coale and Demeny (1983, 47, 52). The log-change of the sex ratio follows from the ${}_1L_0$ values by sex. The same pattern is evident in a wide range of other life tables, including the UN model life tables.

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

[8] Important contributions include Catalano et. al. (2003, 2005, 2005, 2009, 2013) and Almond and Edlund (2007); see more references there. More recently, Morse and Luke (2021) have argued that population-level differences in sex ratios at birth come from maternal health and fetal loss.

[9] Recent work in health economics has followed this logic, using sex ratios at birth as an indicator of fetal loss. See, for example, Sanders and Stoecker (2015), Valente (2015) and Guimbeau et. al. (2020).

2.3 Sex Discrimination

The most important limitation in using sex ratios as an indicator of infant mortality is the role of sex discrimination against girls. Female infanticide and mortal neglect of girls reverse the biological advantage of females, pushing sex ratios towards boys.[10] Because of this, populations with high levels of infant mortality can have healthy-looking sex ratios if sex discrimination is enough to offset the biological mortality advantage of infant girls. For example, the 1891 Census of India reports almost 5% more boys than girls under two-years of age in the province of Panjáb—a value consistent with a healthy population. However, looking across ages, the Panjáb sex-ratio data suggest a severe case of missing girls. Among those under the age of five, we observe 8% more boys than girls, and among five-to-ten year-olds, 18% more boys than girls. We do not know the set of sex-selective mechanisms behind these patterns, but the Panjáb in 1891 is a striking example of sex ratios which reflect sex-discrimination rather than maternal-infant health. It is important to check sex ratios at older ages, as, in general, sex-discrimination is evident in a pattern of increasingly male-biased sex ratios after infancy.[11] On the other hand, because societal gender discrimination operates against girls, not boys, it follows that female-biased sex ratios are an unambiguous sign of high infant mortality and maternal distress.

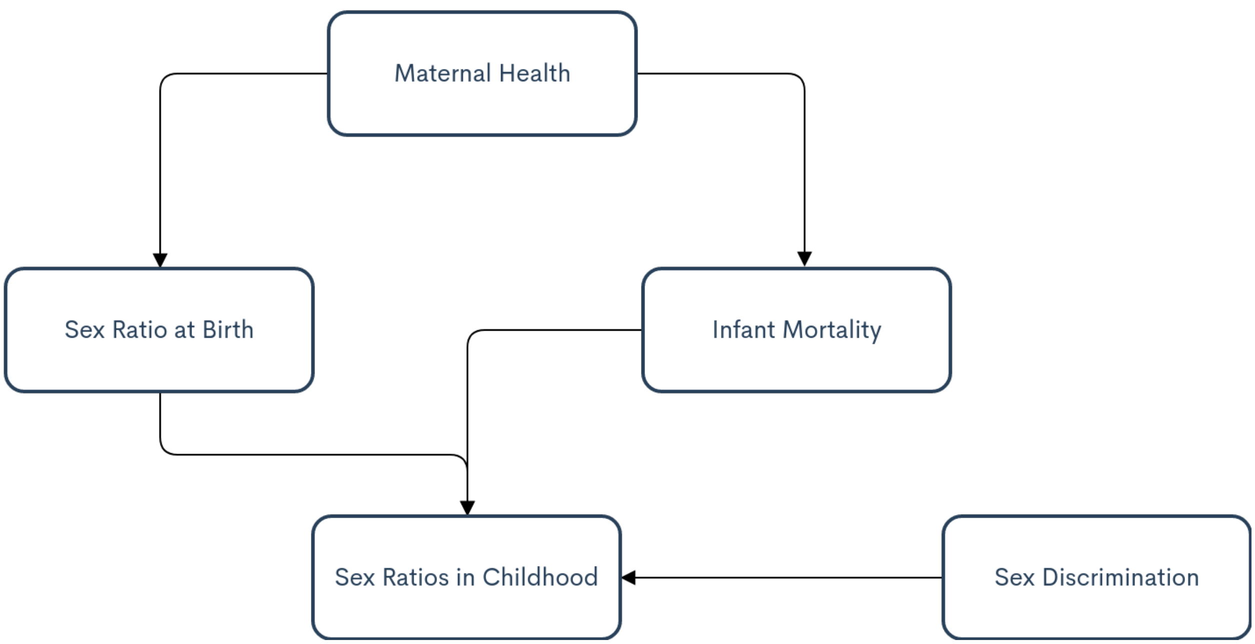

Figure 1: A simple schema of child sex ratios. Maternal health and infant mortality are closely connected, and both determine sex ratios. Sex discrimination acts as a confounding factor.

3 Results

This paper aims to show that sex-ratio data can be used to infer the level of infant mortality in a population. We assemble data on infant mortality and population by age and sex from reliable

[10] For description of the mechanisms and evidence of sex discrimination, see, among many possibilities, the work of Das Gupta (1987).

[11] For example, see the work of Beltrán-Tapia and Raftakis (2020) on modern Greece.

historical sources, comprising most of Europe, North America, and the non-indigenous populations of Australia, New Zealand and South Africa. Starting points for the data span a large range (from 1755 for Sweden to 1920 for Spain), reflecting the uneven development of vital statistics.[12] Because sex ratios provide little information in populations with low levels of infant mortality, and in order to avoid the role of ultrasound in sex-ratio patterns (Campbell 2013), we end our series in the early 1970s. Including data from the 1970s onward would not substantially alter our results. We draw primarily on official sources for vital statistics and census data—see Section 7.3 for more detail.

For each country, we have infant mortality by year and population by age and sex for a subset of years (typically every five or ten years). Within a given year, we take the number of girls under the age of five and divide by the number of boys to obtain the under-five sex ratio: $log(\frac{girls}{boys})$. This under-five sex ratio is paired with the rolling mean of infant mortality, including the year of the population count and the previous four.

Infant mortality and the under-five sex ratio go hand in hand, with an R^2 of 81%. We plot these data in Figure 2. Populations with higher infant mortality have relatively more girls than those with lower infant mortality, and vice versa. With such a close relationship between the two variables, we can infer patterns of infant mortality from sex-ratio data. We fit a median bivariate regression,[13] with infant mortality as the dependent variable and the under-five sex ratio as the independent variable, and obtain: $IMR = 191 + 3250 * log(\frac{girls}{boys})$. We plot this line in Figure 2.

Within our sample, infant mortality and under-five sex ratios are closely related. However, in many historical and contemporary populations we do not have infant mortality data. In such cases, we could use sex-ratio data to *infer* infant mortality. In order to quantify the uncertainty in such inferences, we use out-of-sample prediction.[14] 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. The errors from these predictions (the actual infant mortality minus the predicted) are distributed around zero, with the 5th and 95th percentiles around 45 points overestimation and underestimation respectively. We plot the corresponding 90% prediction interval in Figure 2. See Section 7.2 for further plots of the prediction errors.

[12] The starting points for the data are: Australia (1881–1971), Austria (1869–1971), Belgium (1846–1970), Switzerland (1880–1970), Denmark (1840–1970), Spain (1920–1970), Finland (1885–1970), France (1901–1968), Scotland (1861–1970), England and Wales (1851–1970), Italy (1911–1971), New Zealand (1867–1971), the Netherlands (1859–1970), Norway (1890–1970), Germany (1880–1971), South Africa (1918–1921), and Sweden (1755–1970). For the USA, we have Massachusetts (1860–1970), a subset of states (1920–1930), and the USA as a whole (1940-1970). See Section 7.3 for more detail.

[13] We use median regression as it is less sensitive to outliers than least-squares (Kroenker and Hallock 2001). From a conceptual standpoint median regression is also desirable: we are interested in the *typical* infant mortality rate given a sex ratio, rather than the mean.

[14] For a classic reference, see Butler and Rothman (1980).

Figure 2: Infant mortality rate by log(girls/boys). The black line is the (median) regression of infant mortality on log(girls/boys). The translucent ribbon is the 90% prediction interval, estimated as described in Section 3. The vertical line is a healthy sex ratio at birth of 5% more boys than girls. Data from Europe and settler colonies.

The precision of our predictions depends on the level of infant mortality. In absolute terms, the prediction interval is broadly constant across the sample, with slight heteroskedasticity with respect to infant mortality. However, in real-world applications, we are often more interested in *relative* error than absolute. We plot the proportional error, $log(\frac{Actual}{Predicted})$, against infant mortality in Figure 3. In relative terms, our predictions grow more precise with greater infant mortality. For populations with infant mortality of less than 30 deaths per 1000, the median relative prediction error (i.e. the median of $|log(\frac{Actual}{Predicted})|$) is 78%. For those with infant mortality between 30 and 100 it is 26%, and for those with infant mortality greater than 100 it is 12%.

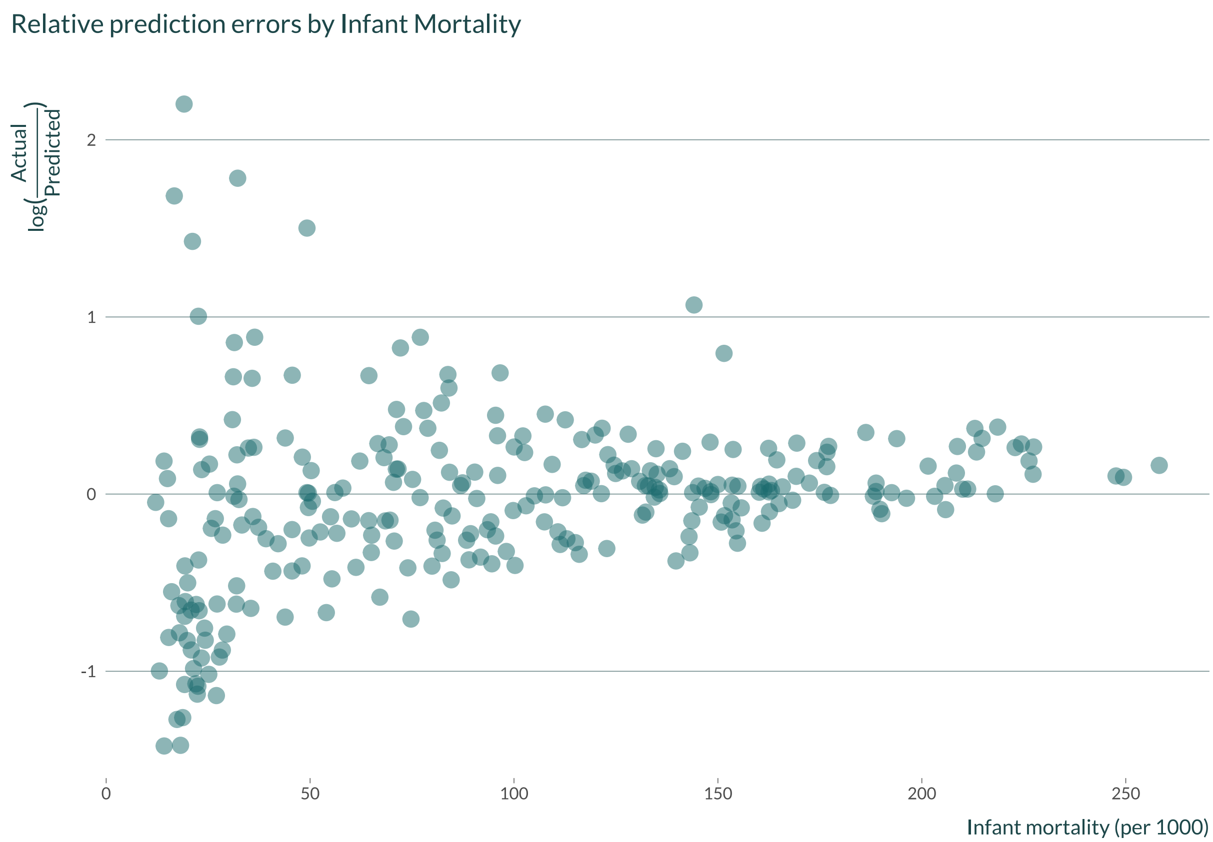

Figure 3: Relative prediction errors, log(Actual/Predicted), by infant mortality. Prediction errors calculated by out-of-sample prediction, as described in Section 3. Data from Europe and settler colonies.

4 Discussion

As seen in the figures above, sex ratios can reveal broad patterns of infant mortality. Infant mortality and the under-five sex ratio are closely related, and the under-five sex ratio can predict infant mortality. These predictions are more precise (in relative terms) for greater levels of infant mortality, making sex ratios particularly useful as indicators in high-mortality populations. As an illustration of the potential of sex ratios to reveal population health, consider the black population of South Africa under the British Empire, for whom we have no direct evidence on infant mortality. In 1911 South Africa, there were 7% more girls than boys under the age of five recorded in the

"Bantu" population.[15] Even allowing for a two-point difference in the sex ratio at birth between white and black populations,[16] the regression in Figure 5 suggests an infant mortality of over 350 deaths per 1000 births.[17] At the same time, the white population of South Africa enjoyed one of the lowest infant mortality rates in the world, around 90, and, accordingly, a sex ratio of 3% more boys than girls. These results demonstrate the tremendous potential of sex ratios to shed new light on infant mortality, particularly in populations ignored by official health statistics. Populations across the British Empire and the Americas had censuses from the late 19th century onward, usually decades before the development of local vital statistics. Sex ratios can shed new light on these populations, which lack adequate records of infant mortality.

However, predictions of infant mortality from sex ratios are, as shown by our data, coarse. To put it simply, within our sample we can use sex ratios to distinguish between infant mortality levels of 100 and 150 deaths per 1000, but not between 120 and 140, nor between 10 and 30. This uncertainty likely comes from differences in maternal health and excess male mortality. Although they are closely connected, infant mortality can vary independent of maternal health, for example due to breastfeeding practices, leading to different sex ratios at birth for populations with the same infant mortality. Furthermore, excess male mortality is not fixed, and depends on the disease environment and socio-cultural matrix into which babies are born (Waldron 1998). Beyond these two factors, there is random variation in the sex ratio at birth. All of these factors introduce imprecision into the attempt to predict infant mortality from sex ratios. However, even with all of this uncertainty, our results demonstrate that the simplest possible model of sex ratios and infant mortality—a bivariate regression—can reveal broad patterns of infant mortality.

Beyond any imprecision, sex ratios are, as discussed above, a fundamentally *one-sided* indicator of maternal-infant health, due to female infanticide and mortal neglect. This issue is apparent in our data. Every population with more girls than boys has infant mortality of at least 150 deaths per 1000 births. However, the association of low infant mortality with male-skewed populations is less universal. In early-20th century Italy, infant mortality was greater than 150, yet there were still 4% more boys than girls. Moreover, as infant mortality decreased in Italy, we see very little change in the relative number of girls and boys. Italy highlights a limitation of our approach and the one-sidedness of sex ratios as an indicator, with male-skewed populations reflecting sex discrimination (or poor data quality) rather than low mortality.[18]

Aside from infant mortality, sex ratios may also prove useful in the measurement of stillbirths and neonatal mortality, which are often poorly recorded in both historical and contemporary populations (Dudfield 1912; Breschi et al. 2012; Gourdon and Rollet 2009; and Gourbin and Masuy-Stroobant 1995; as cited in Pozzi and Fariñas 2015). Perinatal mortality is heavily skewed toward males

[15] The 1911 Census use race classifications of "Bantu", "European or White" and "Mixed and Coloured Races other than Bantu".

[16] See Morse and Luke (2021) for recent analysis. Note that the authors argue convincingly that this difference in the sex ratio at birth reflects maternal health and fetal loss. Therefore, our estimate of infant mortality for the black South African population is conservative with respect to population health as a whole, as we are ignoring this maternal health effect.

[17] Marco-Gracia and Fourie (2021) argue that Black South African sex ratios reflect gender discrimination against young boys; however, we view poor living conditions as a more plausible explanation.

[18] We note that the earlier Italian data, from 1871–1901, shows a similar pattern of male-biased sex ratios. We take our starting date as 1906 for Italy, following the Human Mortality Databases note (2021, 5) that "the [Italian] data prior to 1906 should be used with extra caution due to problems of quality". Regional analysis, see Section 7.1, shows that while central and northern Italy fits into the expected pattern of sex ratios and infant mortality, southern Italy is extremely male-biased. E.g. in 1901, Sicily had over 7% more boys than girls under the age of five and *14%* more girls than boys under the age of 6 weeks).

(Waldron 1998, 64), meaning that changes in the rate of stillbirth and neonatal mortality should be reflected in sex ratios. Sex ratio analysis could also complement existing methods of indirect mortality estimation, e.g. the Brass method or model life tables, which rely on survey questions about child survival and must make strong assumptions about the age-path of mortality.[19] Using sex ratios to estimate infant mortality relies on different assumptions than these modeling techniques, meaning they could be used to corroborate each other.

Sex ratios are promising indicators of infant mortality, but have several practical limitations. First, large sample sizes (tens of thousands) are needed for a sex ratio to have reasonably sized confidence intervals.[20] Second, small changes in infant mortality may prove difficult to discern using sex ratios due to random variation in the sex ratio at birth. Third, sex ratios are also vulnerable to sex-biased age heaping (e.g. differential age reporting of girls and boys). In our work on the US census (McDevitt-Irwin and Irwin, 2021), such sex-biased age-heaping can be observed at ages 0 and 1. To combat these concerns, we focus our attention on the sex ratios for ages 0 to 4, which reduces both age heaping and sampling variation.

5 Conclusion

Child sex ratios can reveal *broad* patterns of infant mortality. Such revelations, though coarse, could be extremely useful in shedding light on hitherto under-studied populations. Wherever we have censuses, but lack satisfactory vital statistics, sex ratios could provide valuable evidence on population health. In this essay, we focus on infant mortality, as our simple, bi-variate approach cannot say much about maternal health. However, health during pregnancy is, for the infant, a critical phase of physical, emotional and cognitive development,[21] making maternal health a central part of population health. Sex-ratio data could provide new evidence on women's health in historical populations.

The concern of missing girls lurks behind any claims to use sex ratios as an indicator of infant mortality. Indeed, sex discrimination against girls will always bias estimates of infant mortality from sex ratios. However, this bias goes in only one direction—*against girls*—meaning that sex ratios provide a *strong* test of *high* infant mortality. When we see 7% more girls than boys, as in the black population of South Africa in 1911, we must conclude that infant mortality was extremely high. Sex ratios have the potential to reveal patterns of infant mortality and maternal health in a wide range of hitherto under-studied populations.

[19] See Hill (1991, 371-374) for a discussion of such methods.

[20] Recalling that a low-mortality population should have about 5% more boys than girls under the age of five, consider an observed ratio of 3% more boys than girls. With 10,000 children the 90% confidence interval is 6 percentage points.; with 50,000 children the 90% CI shrinks to about 3 percentage points (treating the sex-proportion as a binomial random variable – see e.g. Visaria 1967, 133). In the smaller population, the female-biased sex ratio is inconclusive as evidence of poor maternal-infant health. In our own data, Iceland illustrates the limits of sex ratios in small populations. With a population of fewer than 10,000 children for most of the relevant period, the Icelandic data is extremely noisy, and we exclude it from our analysis. The issues of sample size, power and noise have previously been discussed in the context of sex ratios: e.g. Gelman and Weakliem (2009).

[21] See Barker (1995) for a foundational work, and Almond and Currie (2011) for a review of this literature in the social sciences.

6 References

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 Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25 (3): 153–72.

Almond, Douglas, and Lena Edlund. 2007. "TriversWillard at Birth and One Year: Evidence from US Natality Data 19832001." *Proceedings of the Royal Society B: Biological Sciences* 274 (1624): 2491–6. https://doi.org/10.1098/rspb.2007.0524.

Arbuthnot, John. 1710. "II. An Argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes. By Dr. John Arbuthnott, Physitian in Ordinary to Her Majesty, and Fellow of the College of Physitians and the Royal Society." *Philosophical Transactions of the Royal Society of London* 27 (328): 186–90.

Beltrán Tapia, Francisco J., and Domingo Gallego-Martínez. 2017. "Where Are the Missing Girls? Gender Discrimination in 19th-Century Spain." *Explorations in Economic History* 66 (October): 117–26. https://doi.org/10.1016/j.eeh.2017.08.004.

———. 2020. "What Explains the Missing Girls in Nineteenth-Century Spain?" *The Economic History Review* 73 (1): 59–77. https://doi.org/10.1111/ehr.12772.

Beltrán Tapia, Francisco J., Siegfried Gruber, Bartosz Ogórek, and Mikolaj Szoltysek. 2021. "Inferring 'Missing Girls' from Child Sex Ratios in European Historical Census Data: A Conservative Approach."

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

Breschi, Marco, Massimo Esposito, Stanislao Mazzoni, and Lucia Pozzi. 2012. "The Sardinian Experience of the Lowest Italian Infant Mortality at the Turn of the Twentieth Century. True or False Empirical Evidence?" In *Annales de Démographie Historique*, 63–94. 1. Belin.

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.

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

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

Catalano, Ralph A., Katherine Saxton, Tim Bruckner, Sidra Goldman, and Elizabeth Anderson. 2009. "A Sex-Specific Test of Selection in Utero." *Journal of Theoretical Biology* 257 (3): 475–79.

Catalano, Ralph, Tim Bruckner, Elizabeth Anderson, and Jeffrey B. Gould. 2005. "Fetal Death Sex Ratios: A Test of the Economic Stress Hypothesis." *International Journal of Epidemiology* 34 (4): 944–48.

Catalano, Ralph, Tim Bruckner, Jeff Gould, Brenda Eskenazi, and Elizabeth Anderson. 2005. "Sex Ratios in California Following the Terrorist Attacks of September 11, 2001." *Human Reproduction*

20 (5): 1221–7.

Catalano, R., T. Yorifuji, and I. Kawachi. 2013. "Natural Selection in Utero: Evidence from the Great East Japan Earthquake." *American Journal of Human Biology* 25 (4): 555–59.

Coale, Ansley J. 1989. "Demographic Transition." In *Social Economics*, 16–23. Springer.

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

"Completeness of Birth Registration (%) | Data." n.d. https://data.worldbank.org/indicator/SP.REG.BRTH.ZS.

Dana A. Glei, Mila Andreeva, Tim Riffe, Gabriel Borges, and Denys Dukhovnov. 2021. "About Mortality Data for Italy." Human Mortality Database.

Darling, Jonathan C., Panagiotis D. Bamidis, Janice Burberry, and Mary C. J. Rudolf. 2020. "The First Thousand Days: Early, Integrated and Evidence-Based Approaches to Improving Child Health: Coming to a Population Near You?" *Archives of Disease in Childhood* 105 (9): 837–41. https://doi.org/10.1136/archdischild-2019-316929.

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

Dudfield, Reginald. 1912. "Still-Births in Relation to Infantile Mortality." *Journal of the Royal Statistical Society* 76 (1): 1–26.

Gelman, Andrew, and David Weakliem. 2009. "Of Beauty, Sex and Power: Too Little Attention Has Been Paid to the Statistical Challenges in Estimating Small Effects." *American Scientist* 97 (4): 310–16.

Gourbin, G., and Godelieve Masuy-Stroobant. 1995. "Registration of Vital Data: Are Live Births and Stillbirths Comparable All over Europe?" *Bulletin of the World Health Organization* 73 (4): 449.

Gourdon, Vincent, and Catherine Rollet. 2009. "Les Mort-Nés à Paris Au XIXe Siècle: Enjeux Sociaux, Juridiques et Médicaux d'une Catégorie Statistique." *Population* 64 (4): 687–722.

Grech, Victor. 2020. "The Sex Ratio at Birth Aspects." *Early Human Development* 140 (January): 104857.

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." National Bureau of Economic Research.

Gupta, Monica Das. 1987. "Selective Discrimination Against Female Children in Rural Punjab, India." *Population and Development Review*, 77–100.

Hill, Kenneth. 1991. "Approaches to the Measurement of Childhood Mortality: A Comparative Review." *Population Index* 57 (3): 368–82. https://doi.org/10.2307/3643873.

Hollingshaus, Mike, Rebecca Utz, Ryan Schacht, and Ken R. Smith. 2019. "Sex Ratios and Life Tables: Historical Demography of the Age at Which Women Outnumber Men in Seven Countries, 18502016." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52 (4): 244–53.

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

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

Klasen, Stephan, and Claudia Wink. 2002. "A Turning Point in Gender Bias in Mortality? An Update on the Number of Missing Women." *Population and Development Review* 28 (2): 285–312.

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.

Marco-Gracia, Francisco J., and Johan Fourie. 2021. "The Missing Boys: Understanding the Unbalanced Sex Ratio in South Africa, 18942011." *Economic History of Developing Regions*, 1–19.

McDevitt-Irwin, Jesse, and Irwin, James R. 2021. "Infant Sex Ratios and Infant Mortality in North America: 1850-1920." Unpublished Manuscript.

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

OECD, and WHO. 2020. *Infant Mortality.* Doi:https://doi.org/10.1787/f7b3e551-en.

Pozzi, Lucia, and Diego Ramiro Fariñas. 2015. "Infant and Child Mortality in the Past." In *Annales de Démographie Historique*, 55–75. 1. Belin.

Reidpath, D. D., and P. Allotey. 2003. "Infant Mortality Rate as an Indicator of Population Health." *Journal of Epidemiology & Community Health* 57 (5): 344–46. https://doi.org/10.1136/jech.57.5.344.

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.

Tavassoli, Nahid, Hamid Noghanibehambari, Farzaneh Noghani, and Mostafa Toranji. 2021. "Upswing in Industrial Activity and Infant Mortality During Late 19th Century US." Papers 2101.02590. arXiv.org.

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

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?*, 64–83. 155. New York: United Nations.

7 Appendix

7.1 Italian Regions

When infant mortality and sex ratios are broken out by region, it is clear that the southern provinces are driving the anomalous results in Italy. Figure 4 plots sex ratios and infant mortality in Italy 1901–1971. For Central and Northern Italy (everything north of Abruzzo and Campania), the relationship between sex ratios and infant mortality fits well with the rest of our data. Southern Italy, on the other hand, is an outlier. In southern Italy, sex ratios are consistently male biased. Despite infant mortality approaching 200 deaths per 1000, there are 5–6% more boys than girls under the age of five in Southern Italy in 1901, 1911 and 1921. This must either be due to sex discrimination or data quality issues.

Figure 4: Infant mortality rate by log(girls/boys). Southern vs Northern Italy. The lines represent median regressions. The vertical line represents a healthy sex ratio at birth, with 5% more boys than girls. Data from ISTAT.

7.2 Additional figures

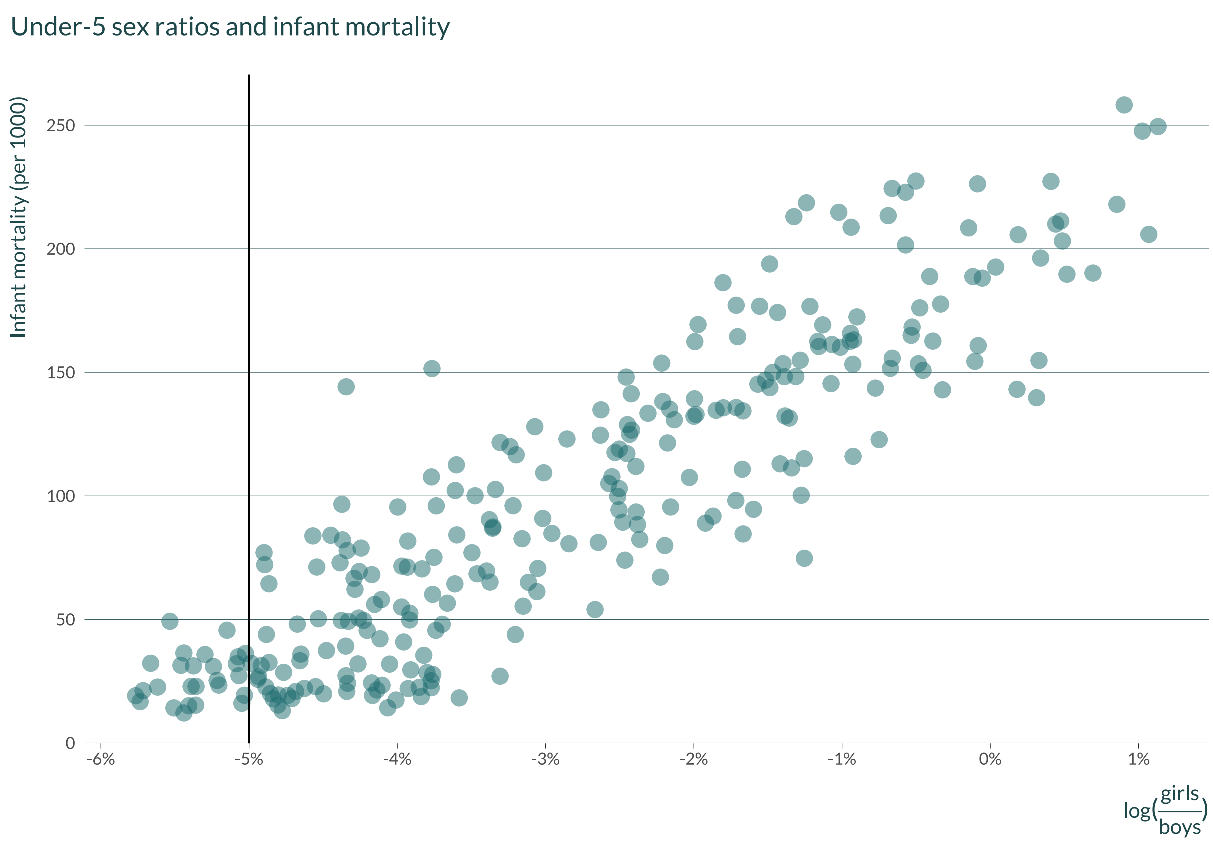

Figure 5: Infant mortality rate by log(girls/boys). The vertical line represents a healthy sex ratio at birth of 5% more boys than girls. Data from Europe and settler colonies.

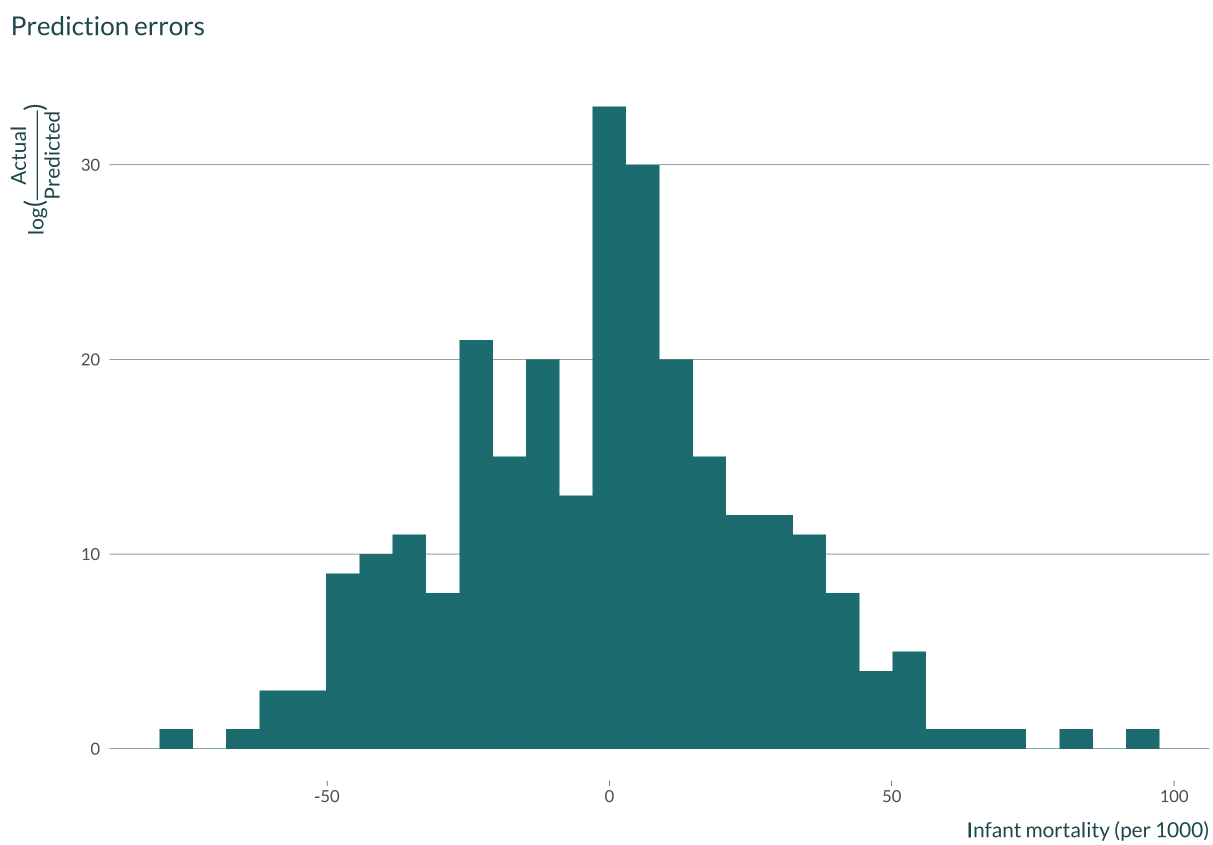

Figure 6: Prediction errors from bivariate median regression of infant mortality on the under-five sex ratio. Data from Europe and settler colonies.

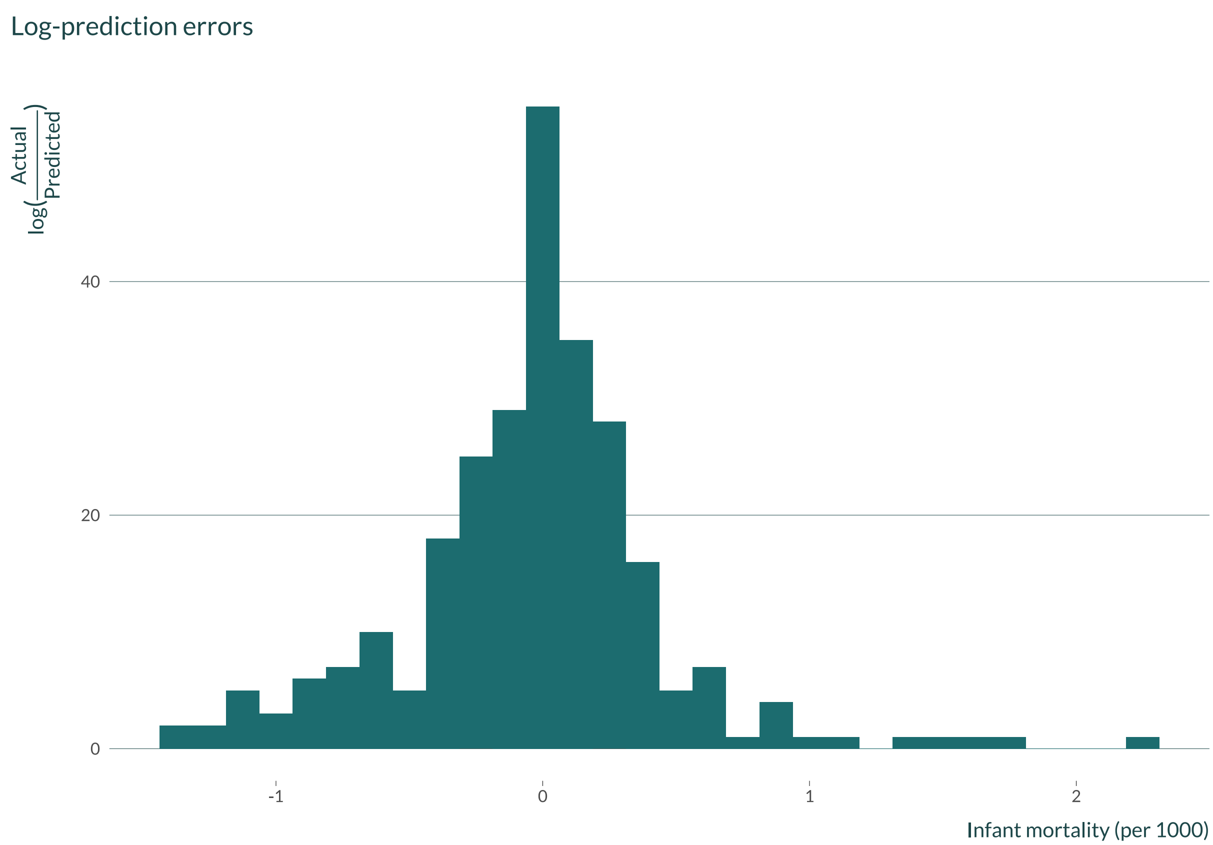

Figure 7: Relative prediction errors, log(Actual/Predicted), from bivariate median regression of infant mortality on the under-five sex ratio. Data from Europe and settler colonies.

7.3 Data Sources

The *Human Mortality Database* (HMD) provides appropriate data for infant mortality rates and sex ratios for many historical populations. We expand our geographic scope by drawing on vital statistics and census data from various official sources. In many cases, the data are available from *International Historical Statistics* (IHS 2013).

In general, we calculate sex-ratio values from official population counts by sex and age; for registry-based sex-ratios, we take value at five-year intervals. For infant mortality rates, we rely on official vital statistics except when demographic scholarship offers better estimates. Specific sources and methods by country follow.

Australia (1880-1971)

Infant mortality rates:

1870-1901: Series MFM 154 in Peter McDonald et al., "Marriage, Fertility, and Mortality," Chapter 3 of Vamplew (1987).

1901-1971: Australian Bureau of Statistics, Australian Historical Population Statistics, 3105.0.65.001, Table 5.4 "Infant mortality rates ... 1901 onwards" (released 2019 April 18).

Under-5 populations by sex (census values, non-aboriginal populations):

1881 & 1891: census values reported in J. C. Caldwell, "Population", Chapter 2 of Vamplew (1987).

1901 & 1911: *Census of the Commonwealth of Australia ... April, 1911.* Volume II, pp. 10, 11.

1921, 1933, 1947, 1954, 1961, 1966: *Census of Population and Housing, 30 June 1966, Commonwealth of Australia. Volume 1. Population: Single Characteristics,* pp. 10-11.

1971: *Census of Population and Housing, 30 June 1971, Commonwealth of Australia, Bulletin 1. Summary of Population, Part 9 Australia,* p. 1; *Bulletin 9. The Aboriginal Population,* p. 1.

Until 1971, the Australian census tables excluded the Aboriginal population.

Australian Bureau of Statistics, Australian Historical Population Statistics, 3105.0.65.001, Table 2.1 "Population, age and sex, Australia, 30 June, 1901 onwards" (released 2019 April 18). Vamplew, Wray, ed. 1987. *Australians: Historical Statistics.* Broadway, New South Wales: Fairfax, Syme & Weldon.

Austria (1869-1971)

Infant Mortality Rates. 1865-1971: *IHS* (2013: 3577, 3580, 3583), Series A7.

Under-5 populations by sex. 1869 & 1910: *IHS* (2013:3440), Series A2 (Austrian provinces of the Hapsburg Empire). The values here are rounded to the nearest thousand (we would prefer unrounded data, but were unable to locate the data in official sources). 1880, 1890, 1900: *Österreichisches statistisches Handbuch ... 1886,* p. 3; *1893,* p. 6; 1909, p. 7. 1923, 1934, 1951, 1961 Statistik Austria, *Statistisches Jahrbuch 2010,* p. 45 (2.08 Bevölkerung 1869 bis 2001 nach fünfjährigen Altersgruppen und Geschlecht (Population 1869 to 2001 by five-year age groups and sex)).

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

Belgium:

1846-1970 from the HMD.

Denmark:

1840-1970 from the HMD.

England and Wales:

1851-1970 from the HMD.

Finland:

1885-1970 from the HMD.

France:

1901-1968 from the HMD.

Germany (1876-1933)

Infant Mortality Rates: IHS (2013: 3577, 3580), Series A7.

Under-5 populations by sex.

1880: *Statistisches Jahrbuch 1883*, p. 10; 1890: *Statistisches Jahrbuch 1896*, p. 5; 1900: *Statistisches Jahrbuch 1903*, p.6; 1910: *Statistisches Jahrbuch 1919,* pp. 6-7; 1925: *Statistisches Jahrbuch 1929*, p. 14; 1933: *Statistisches Jahrbuch 1939*, p. 14. IHS reports these data rounded to the nearest thousand (Series A2, p. 3454). For 1933, we use the census values, dated 16 June 1933; the IHS value for 1933 matches the estimates for Dec. 31, 1933, found in *Statistisches Jahrbuch 1936*, p. 12.

West Germany (1951-1970)

> HMD provides under-5 populations by sex for 1956, 1961, and 1970, and infant mortality rates for 1951-56, 1957-61, 1966-70)

East Germany (1959-1971)

> HMD provides under-5 populations by sex for 1964 and 1971, and infant mortality rates for 1959-64 and 1966-71.

Italy (1911-1971)

Infant mortality rates: ISTAT

Under-5 Population by Sex:

1911-1949: Cox, P. R. (1958). The aging of populations and its economic and social implications. UN. Pages 142-143.

1951-1971: taken from ISTAT

Sources for Italian regional data (all available online from the Istat Digital Library).

IMR Tendenze evolutive della mortalità Infantile in Italia. Annali di Statistica, Serie 8 Volume 29. Roma: 1975.

U5 populations by sex for Italy's regions are found in the following published census volumes.

1901 Numero delle famiglie e numero degli abitanti classificati secondo la qualita della dimora, il luogo di nascita, il sesso, l'età, lo stato civile e l'istruzione. Vol. 2 of Censimento della popolazione del Regno d'Italia al 10 febbraio 1901. Roma: Tipografia Nazionale, 1903.

1911 Popolazione presente classificata per sesso, età, stato civile ed istruzione. Vol. 2 of Censimento della popolazione del Regno d'Italia al 10 giugno 1911. Roma: Tipografia Nazionale, 1914.

1921 Risultati sommari del censimento della popolazione eseguito il 1. dicembre 1921. Italia : Direzione generale della statistica : Ufficio del censimento. Roma: Societa anonima G. Scotti, 1925-26. Volumes 1-18 (each volume covers a region).

1936 Regno : popolazione, territorio, famiglie, convivenze, sesso, stato civile, età, stranieri. Parte Seconda - Tavole. Vol. 3 of VIII Censimento generale della popolazione : 21 aprile 1936, Istituto centrale di statistica del Regno d'Italia. Roma: Tipograpfia Failli, 1937.

1951 Sesso, età, stato civile, luogo di nascita. Vol. 3 of IX Censimento generale della popolazione : 4 novembre 1951, Istituto centrale di statistica. Roma: Soc. Abete, 1956.

1961 Sesso, età, stato civile, luogo di nascita. Vol 5 of X Censimento generale della popolazione : 15 ottobre 1961, Istituto centrale di statistica. Roma: 1968.

1971 Sesso, età, stato civile. Vol 5 of XI Censimento generale della popolazione : 24 ottobre 1971, Istituto centrale di statistica. Roma: 1974.

New Zealand (1864-1971)

Under-5 populations by sex

1864, 1867, 1871, 1874, 1878, 1881: Results of a Census of the Colony of New Zealand .... Chapter 28, Table 1 ("Showing the Increases of Persons ... at different Ages ... the various censuses ... 1864 ... 1881).

1886, 1891, 1896, 1901, 1906, 1911, 1916: Results of a census of the Dominion of New Zealand ... 1916, Part II Ages, page 1.

1936, 1945, 1951, 1956, 1961, 1966, 1971: Stats NZ Spreadsheet A1.6 (citing Bloomfied (1984), "Census Reports: Table II.6. Age Groups . . . 1874-1976"). Available in the "discontinued series" of Statistics NZ Infoshare, Census of Population and Dwellings, Total Pop - Sex by Single Years of Age. 1874-1945 (excluding Maori population; 1951-1971 total population).

Netherlands:

1859-1970 from the HMD.

Norway (1890-1970)

Census counts of under-5 populations by sex start with the year 1845, but reliable IMR data for Norway start with 1876.[22] 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

[22] Census U5 counts are available for 1845, 1855, 1865, 1875, but the 1890 count is the first for which we also have a reliable 5-year average infant mortality rate. 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).

correspondingly understated.[23] Although some early publications from Statistics Norway report IMR data from before 1876, their *Historical Statistics* of 1978, 1994, and 2000 present 5-year average values of IMR starting with 1876. In our view, that corroborates our conclusion that 1876 marks the start of reliable IMR data for Norway.

IMR: annual values 1886-1970, as published in IHS, corroborated with Statistics Norway data.

U5 census populations by sex: HMD (1890-1970).

Statistics Norway (1910), *Norges folkemængde fordelt paa de enkelte aldersaar 1846-1901.* Norges Officielle Statistik V. 113. Kristiania (Oslo), H. Aschehoug & Co.

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).

Scotland:

1861-1970 from the HMD.

South Africa (1918, 1921)

IMR 1913-1921: *International Historical Statistics 1750-2010: Africa, Asia and Oceania (Volume 1),* Series A7 Africa, p. 1 (DOI 10.1057/9781137305688).

U5 populations by sex, 1918: *Official Yearbook of the Union ... 1922*, pp. 158-59; 1921: *Official Yearbook of the Union ... 1925*, p. 868.

Spain:

1920-1970 from the HMD.

Sweden:

1755-1970 from the HMD.

Switzerland (1880-1970)

Infant mortality rates (1875-1970): calculated from Historical Statistics of Switzerland HSSO, 2012. Tab. C.41 (deaths by age) and Historical Statistics of Switzerland HSSO, 2012. Tab. C.5a (live births). These IMRs are corroborated with IHS (2013:3578,3582), Series A7.

Under-5 populations by sex: Historical Statistics of Switzerland HSSO, 2012. Tab. B.8a .

United States (1920-1970)

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 ¼ through the year).

[23] 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 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.

IMR (1915-1920, 1925-1930) from *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 (online from NBER). Birth counts are used to aggregate state-level rates; the births data are from the annually published *Mortality Statistics* (online from the National Center for Health Statistics).

For 1915-1920, the states included 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.

1940-1970: HMD provides both IMR and under-5 population by sex.

Massachusetts (1860-1970)

IMR 1856-1970: HSUS (2006) Series Ab928 (Infant mortality rate for Massachusetts: 1851–1998).

U5 populations by sex,

1860 to 1970 at 10-year intervals: US Census Office publications and data-sets.

1865-1915 at 10-year intervals, and 1955: State censuses of Massachusetts.