Historical Methods: A Journal of Quantitative and Interdisciplinary History

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/vhim20

Inferring "missing girls" from child sex ratios in historical census data

Mikołaj Szołtysek, Bartosz Ogórek, Siegfried Gruber & Francisco J. Beltrán Tapia

To cite this article: Mikołaj Szołtysek, Bartosz Ogórek, Siegfried Gruber & Francisco J. Beltrán Tapia (2022) Inferring "missing girls" from child sex ratios in historical census data, Historical Methods: A Journal of Quantitative and Interdisciplinary History, 55:2, 98-121, DOI: 10.1080/01615440.2021.2014377

To link to this article: https://doi.org/10.1080/01615440.2021.2014377

Published online: 18 Feb 2022.

HISTORICAL METHODS: A JOURNAL OF QUANTITATIVE AND INTERDISCIPLINARY HISTORY
2022, VOL. 55, NO. 2, 98–121
https://doi.org/10.1080/01615440.2021.2014377

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Inferring "missing girls" from child sex ratios in historical census data

Mikołaj Szołtysek[a], Bartosz Ogórek[b], Siegfried Gruber[c], and Francisco J. Beltrán Tapia[d]

[a]Institute of Sociological Sciences, The Cardinal Wyszyński University, Warsaw, Poland/Hungarian Demographic Research Institute, Budapest, Hungary; [b]Institute of History and Archival Science, Pedagogical University of Cracow, Cracow, Poland; [c]Institute of History, University of Graz, Graz, Austria; [d]Department of Modern History and Society, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT
The topic of "missing girls" in historical Europe has not only been mostly neglected, but previous research addressing this issue usually took the available information too lightly, either rejecting or accepting the claims that there was discrimination against female children, without assessing the possibility that the observed child sex ratios could be attributable to chance, mortality differentials, or registration quality. This article contributes to this discussion by (1) using a novel dataset of historical child sex ratios that covers a large part of the European geography between 1700 and 1926; and (2) explicitly considering the effects of random variability, demographic variation, and faulty enumeration in the analysis. Our results provide evidence that some of these European populations had child sex ratios well above the levels usually considered "natural". Although part of this variation is indeed shown to be due to random noise and structural features related to infant mortality differentials and census quality, some of the observed sex ratios are too high to be attributed solely to these proximate factors. Thus, these findings suggest that there are behavioural explanations for some of the unbalanced sex ratios observed in our data.

KEYWORDS
Sex ratios; gender discrimination; infanticide; infant mortality; census microdata

Introduction and background

Son preference resulting in discrimination against daughters, either via sex-selective abortion, female infanticide and/or mortal neglect of girls during infancy and childhood, have generated highly unbalanced sex ratios in South and East Asia (e.g., Sen 1990; Das Gupta et al. 2003; Jayachandran 2015; Guilmoto 2018). However, the question of whether similar forms of gender discrimination might have also occurred among earlier generations of Europeans has received surprisingly little attention. One reason for this imbalance is related to the widespread narrative arguing that household formation patterns and religious prescriptions left little room for gender-specific forms of mortal neglect in the European past, at least in the last few centuries (Lynch 2011; also Derosas and Tsuya 2010).[1]

However, perhaps even more importantly, assessing whether girls indeed went "missing" in European history is a very challenging task. Direct evidence of infanticide or the mortal neglect of female offspring in historical Europe is scarce at best. Moreover, efforts to indirectly infer these phenomena from sex ratio data (the number of males divided by the number of females) are also riddled with considerable problems.[2] The usual method employed to deduce discriminatory practices leading to excess female mortality in infancy and childhood compares the actual age-specific sex ratio with the "expected" one, that is, the ratio that would result from equal treatment of the sexes in the distribution of survival-related ends (Klasen 1994; Klasen and Wink 2003). However, the efforts to set a nondiscriminatory standard against which particular historical conditions can be evaluated have been speculative at best, as we simply do not know how sex ratios should have looked like in the past. Although various "natural" sex ratio standards have been proposed for contemporary human populations (Visaria 1967; Coale 1991; Klasen and Wink 2003; also Johansson and Nygren 1991), attempts to simply apply them to historical conditions have been superficial, if not misleading (James 1997), especially in

CONTACT Mikołaj Szołtysek m.szoltysek@uksw.edu.pl The Cardinal Wyszyński University, Warsaw, Poland.
Supplemental data for this article is available online at http://dx.doi.org/10.1080/01615440.2021.2014377

large-scale comparative projects.[3] This is because the combined effect of random variation, overall mortality conditions, and/or the sex-selective undercounting of children may result in abnormally high (or low) sex ratios even in the absence of any preferential treatment by parents. Given the suspicion that any historical statistics may be unreliable, and the demographic variability of historical Europe, scholars must be able to show that male-skewed sex ratios did not arise artificially due to mere chance, mortality differentials, or faulty enumeration before they give credence to more behavioral explanations.[4]

This conundrum has long been recognized. As early as the beginning of the 20th century, Gini (1908) pointed out the risk of considering random fluctuations in sex ratios at birth as significant patterns arising from demographic and environmental determinants (similarly Henry and Blum 1988; Delille 1974; or Fellman 2015). Likewise, discussing skewed sex ratios in the Florentino census of 1427, Herlihy and Klapisch-Zuber (1985, 131–135) wondered whether they were indicative of gender differentials in mortality arising from indifference or negligence toward female offspring, or mere reflections of defects in the data collection and registration (cf. Bender 2011). Similarly, Ring (1979) advised strongly against making careless inferences of infanticide from medieval sex ratios alone unless all possibilities of errors in the data could be eliminated. More recently, it has been shown that unconditional generalizations based on the sex ratios of children can be extremely risky, as these ratios are influenced by the underlying levels of infant mortality (Beltrán Tapia and Gallego-Martínez 2017; Beltrán Tapia 2019). Finally, other scholars have provided evidence that the registration of the sexes may be affected in various ways in different census types (Szołtysek 2015, v. 2; also Bender 2011; Emigh, Riley, and Ahmed 2016).

Unfortunately, in the historical sex ratio literature, many scholars are still failing to heed these calls for caution, as they often reach the conclusion that sex ratios were high without considering alternative interpretations of their data. In response to this unsatisfactory situation, we argue that any interpretation of the "missing girl" phenomenon in historical Europe should be based on a thorough understanding of the intricacies of the historical data customarily used to derive sex ratios and their statistical properties, as well as of the underlying demographic features of these populations. In this article, we raise fundamental questions regarding the claims made about the male-skewed ratios in historical Europe that are sometimes observed, and present a methodology that should provide better answers to the following questions: What did child sex ratios look like in historical Europe, and how trustworthy are these values? What proximate factors, other than the willful neglect of female offspring, might explain the elevated sex ratios in historical populations? And, most crucially, can the unusual surplus of male children that is often detected in the data still be observed after controlling for the possible influence of random noise, overall infant mortality, and data quality?[5]

We extend the existing literature in five major directions. First, we expand the geographic and temporal coverage of earlier studies. Previous attempts to map historical sex ratios in Europe have been limited to particular areas or specific case studies (see Reynolds 1979; Bechtold 2001; Hynes 2011; Hanlon 2003, 2016, 2017; Kemkes 2006; Beltrán Tapia and Marco-Gracia 2021; Beltrán Tapia and Raftakis 2021; Marco-Gracia and Beltrán Tapia 2021; also Manfredini, Breschi, and Fornasin 2016; Sandström and Vikström 2015; Johansson 1984; Coleman 1976). Larger-scale comparative data on historical sex ratios have been studied primarily for the late 19th century, and then only for western and southern Europe (Bechtold 2006; Beltrán Tapia 2019; Beltrán Tapia and Gallego-Martínez 2017, 2020; also Charpentier and Gallic 2020).[6] By providing an unprecedented dataset of historical child sex ratios in more than 300 regional populations stretching from Andalusia in the west to Siberia in the east, and from Tromsø in the north to Albania in the south-east, we are in a better position to disentangle the variation in European child sex ratios in the past, and the basic factors conditioning it. Particularly noteworthy is the inclusion in our analysis of multiple eastern and south-eastern European societies that were characterized by the rigid forms of patriarchal bias that are often associated with the "missing girls" phenomenon (Szołtysek et al. 2017; cf. Lynch 2011; Miller 2001; Greenhalgh 2013).

Second, our article accounts explicitly for the uncertainty arising from small sample sizes. While aggregate census tracts provide less noisy estimations due to the law of large numbers, they may conceal as much as they reveal, since local-regional sex ratio imbalances may go undetected (Fossett and Kiecolt 1991, 942). However, data on sex ratios extracted from smaller populations may be fraught with estimation uncertainty (e.g., Visaria 1967; also Guilmoto and Oliveau 2007). As some of our regional data indeed suffer from the "small-N problem", analyses of these data may result in excessive sex ratios simply by chance. Here, we explicitly take these risks into account by considering the underlying random variability arising from differences in sample sizes across locations when assessing their sex ratios. Using this approach, we are able to control for a crucial confounding factor that many previous studies have overlooked, while at the same time providing insights into potentially important local variations across historical Europe.

Third, we take regional infant mortality rates directly into consideration when modeling historical sex ratios. This is critical because, due to the female biological advantage, higher mortality rates translate directly into sex ratios (Klasen and Wink 2003, 269–271; Bhaskar and Gupta 2007; Beltrán Tapia and Gallego-Martínez 2017; Beltrán Tapia 2019). Male vulnerability implies that high-mortality environments take a greater toll on boys than girls, which should lower the child sex ratio. Therefore, mortality differentials affect the expectation of what the "natural" sex ratio should be in a particular population. Likewise, we explicitly take into account the possibility that census quality issues could affect the relative enumeration of boys and girls, and, in turn, bias sex ratios, especially due to the under-reporting of females. We therefore consider different measures of census quality and explore how they relate to the observed sex ratios.

Fourth, the novelty of our study is that we take all these factors into account simultaneously by trying to net out the combined effect of stochastic variability, infant mortality, and data quality when modeling the observed sex ratios. This approach is purposely intended to be a conservative research strategy: i.e., it attempts to explain the variation in child sex ratios through basic features that reflect the statistical and demographic properties and the quality of our data, and that are not necessarily related to the "missing girls" phenomenon.[7] Nonetheless, our results indicate that random noise, the mortality environment, and the quality of the census do not fully explain the variability of child sex ratios. Therefore, we suggest that different discriminatory practices may have unduly increased female mortality early in life in the historic populations of some European regions. This approach constitutes a major advantage over earlier studies that either did not try or were not able to ascertain whether the "abnormal" child sex ratios they found resulted from some confounding factors other than gender-specific discrimination (e.g., Reynolds 1979; Bechtold 2001; Kemkes 2006; Hynes 2011; Hanlon 2016, 2017; cf. discussion in Willigan and Lynch 1982, 83–84; Scalone and Rettaroli 2015; also Bolton 1980; Miller 1989).

Finally, our article contributes to earlier efforts to develop a benchmark against which the observed sex ratios can be evaluated and compared (e.g., Coale 1991; Klasen 1994; Klasen and Wink 2003; Chao et al. 2019; Beltrán Tapia and Gallego-Martínez 2017; Beltrán Tapia 2019; also Miller 1989; Johansson and Nygren 1991). Although we do not know what the expected ("normal") values of the sex ratio should be in each of our populations, we approach this conundrum by estimating what such a value would look like given the underlying infant mortality rate and the quality of the enumeration, as well as the uncertainty arising from the number of observations from which those sex ratios are computed. This is a significant improvement over previous works that either too eagerly accepted the contemporary biological standards of sex ratios as a reference, or gauged possible benchmarks based on model life tables, while ignoring other potentially intervening factors.

The reminder of the article is structured as follows. We begin by reviewing the available data. We then evaluate various features that may explain the observed CSR patterns by focusing on three groups of proximate factors that may "naturally" affect the observed number of boys and girls: random variation, the level of infant mortality, and data quality. In the main empirical section, we run a multivariate regression model in order to simultaneously assess whether the observed child sex ratio in a particular area is still high even after these factors are controlled for. In sections five and six, we show that our baseline results hold up to a battery of robustness tests. In the final section, we further discuss our results.

Materials and methods

Data description

For our analyses, we use the largest publicly-available collection of European historical census microdata, which have been compiled by the North Atlantic Population Project (NAPP; distributed by IPUMS-International; Ruggles et al. 2011) and the Mosaic project (Szołtysek and Gruber 2016; Szołtysek and Poniat 2018). These data are in the form of machine-readable, harmonized samples derived from various kinds of historical census and census-like materials, including full-count national censuses, as well as regional fragments of censuses, church lists of parishioners, tax lists, and local estate inventories; all of which are very similar in terms of their structure,

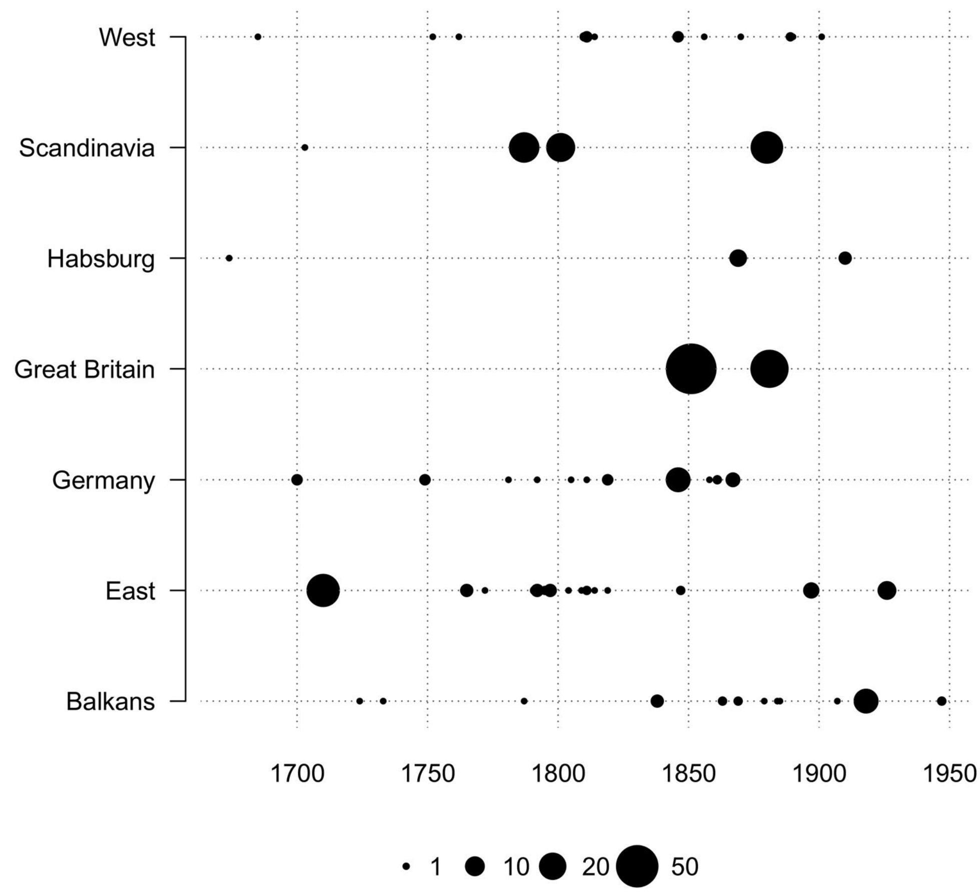

Figure 1. Geographic and temporal variation in the dataset.
Note: The size of the circles indicate the number of regions in each period and region. Seven bigger territorial groupings on the right-side panel of the figure followed major institutional and socioeconomic distinctions across historic Europe. "Great Britain": England, Wales, and Scotland; "Scandinavia": Danish, Swedish, and Norwegian data, as well as Iceland; "Germany": German-dominated areas other than the Habsburg territories; "West": areas west and south-west of Germany; "Habsburg": Austrian, Hungarian, Croatian, as well as Slovakian data; "East": east-central and eastern Europe, including the former Polish-Lithuanian Commonwealth and Russia; "Balkans": areas south and/or east of Croatia and Hungary.
Source. Mosaic/NAPP data. For primary sources of the Mosaic and NAPP data and the full list of items, see section D5 of the Supplemental.

their organization, and the types of information they provide.[8] The samples list all individuals grouped into households (coresident domestic groups) in each settlement or area, and provide information on each individual's sex, age, marital status, and relationship with the head of the household.[9] While the information contained in these listings allows us to derive a large number of family and demographic indicators (e.g., Szołtysek et al. 2020; Szołtysek and Ogórek 2020), we focus here on two basic dimensions of the censuses: age and sex statistics, from which age-specific sex ratios are calculated.

Our approach is situated at the meso-level of comparative analysis, and our units of analysis are "regions". Accordingly, the microdata from 21,559 rural parishes, sub-parishes or communes in the NAPP were aggregated into 156 administrative units that were used in each respective census, and that were considered by the NAPP (generally counties). Likewise, over 4,500 Mosaic locations (settlements, parishes, estates) were agglomerated into 160 regions that correspond either to their respective administrative units (usually also counties), or to geographical clusters in the absence of applicable administrative units.[10] Altogether, we collected information on nearly four million individuals living between roughly 1700 and 1926 in 316 regional populations representing most parts of Europe. Thus, this dataset covers a large share of the variation in historical family and household formation patterns across European populations, as well as across rural and urban contexts, during this period (Szołtysek and Ogórek 2020).

Figure 1 illustrates the distribution of our data across regions and time periods. Of the 316 regional populations, 82 are dated before 1800 (25.9%). These populations are located mainly in eastern and south-eastern Europe, as well as in Scandinavia. The other 18% of the regional populations (N = 57) are from the 1800–1850 period, while the remaining 56% (mostly in Great Britain) date from the post-1851 period. Whereas the pre-1800 locations are geographically clustered, a large share of the populations in the data from north-western Europe come from time periods when the industrial urban revolution was well underway. It is important to note, however, that this data structure stems from the availability of digitized census microdata. Section A of the Supplemental provides further evidence that this data structure does compromise the analyses presented below.

The measure of sex proportions

While much of the literature on sex ratios deals with sex ratios at birth (a flow measure), this article focuses on the child sex ratio (henceforth, CSR), defined as the number of boys aged 0–4 per hundred girls of the same age (a stock measure) (see Klasen and Wink 2003, 265; also Guilmoto and Oliveau 2007). Although the choice of this indicator is dictated by the nature of our data, it has certain advantages. Firstly, as the CSR is a synthetic measure of gender imbalances that incorporates the impact of sex differentials in mortality around birth and during infancy and childhood, it may be able to account for female infanticide, as well as for mortal neglect in the early years of life, including child abandonment (Miller 1989; Agnihotri 1996; Cai and Lavely 2007; Bhaskar and Gupta 2007; den Boer and Hudson 2006).[11] Secondly, this indicator is less subject to the inherent challenges that historical (pre-statistical) societies faced when enumerating live births and infant deaths (Henry 1968; also Chao et al. 2019). Although population censuses were also subject to under-registration, a live toddler was more likely to be counted than a live or deceased infant, an advantage that increased as children grew older. Likewise, the 0–4 age group is not known to be significantly affected by the sex-specific migratory patterns that may deeply alter the sex ratio among older (juvenile) age groups (see below). Finally, by including five annual cohorts, the CSR increases the sample size and reduces the effects of short-term fluctuations and even misplacement (e.g., due to age heaping), and is therefore more robust and statistically stable. In any case, having exact information about the number of boys and girls at each age allows us to compute the sex ratios using other age groups (1–5, 5–9, etc.), which, in turn, enables us to test the robustness of the results achieved with the main measure.

Unconditional data distribution

Figure 2 shows the spatial variation in CSRs that is present in our data. Several features stand out. First, a direct glance at Figure 2 would, at best, suggest a very moderate level of support for the general prevalence of the mortal neglect of girls in our data. The sample mean of CSRs for all 316 regions is slightly above parity (101), and in the majority of our populations, this value is rarely exceeded. However, this average conceals a high degree of internal variation, as almost every major region (perhaps except for England) contains places with very high values of the ratio.[12] Crucially, 73 out of the 316 locations have ratios above 105 boys per 100 girls, a benchmark considered "neutral" by contemporary standards. This suggests that a non-trivial fraction of our data shows evidence of unusually high sex ratios (see also Table 1B in the Supplemental).[13] Although these regions do not form any specific cluster, they tend to congregate somewhat toward the eastern and south-eastern section of our data – especially in parts of European Russia and western Siberia, and in the Balkans.[14] However, various isolated hot spots of similarly male-heavy sex ratios can also be detected in Slovakia and Hungary, in central Poland and Ukraine, as well as in Scotland, Catalonia, southern France, north-western Germany, Austria, and Switzerland.

Overall, the picture provides no clear patterns, since areas with high sex ratios are often contiguous with or relatively close to areas where CSRs are more balanced or even lower. This pattern pertains especially to the Mosaic data, in which the locations exhibit a higher degree of spatial variability (especially in the Russian and German territories). This variability may be attributable to the smaller sample sizes and more uneven quality of the listings in the Mosaic data (cf. Miller 1989, 1231). Moreover, neighboring samples do not always refer to the same historical period, which makes finding spatial patterns more challenging. The 11 Albanian regions are the main exception to this general rule, which is perhaps not surprising given previous evidence that sex-selective discrimination was being practiced in this area well into the 20th century (Grogan 2018).[15] Areas with less extreme, albeit still relatively high child sex ratios (e.g., 102–105), seem to be somewhat more contiguous, as we can observe clusters of similar values throughout Scotland, parts of Denmark and southern Sweden, and a few other more isolated regions. Indeed, the spatial distribution of the CSRs is not entirely random (the global Moran's I index of spatial autocorrelation is 0,27 and statistically significant at the 99%

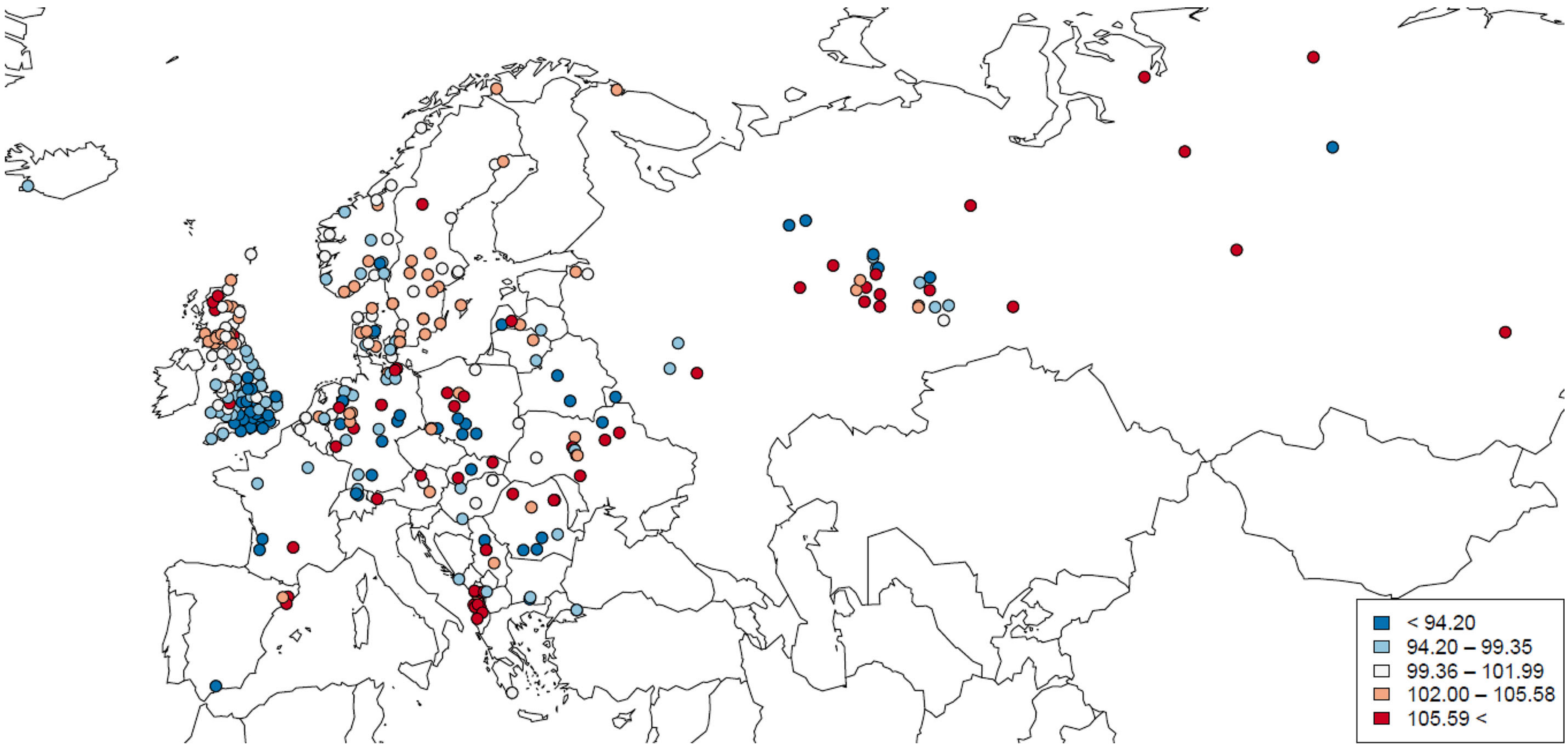

Figure 2. Child sex ratios in the NAPP/Mosaic dataset.
Notes: each point on the map represents the centroid of one Mosaic/NAPP regional population as defined in the text. Whenever applicable, the CSRs are calculated from the weighted numbers of children aged 0-4 (see Section B in the Supplemental).
Source: Mosaic/NAPP data

confidence level), so the question is to what extent this variation can actually be linked to local factors.

Factors affecting child sex ratios

Technically speaking, elevated sex ratios may show up in the data due to some combination of five proximate causes: (1) random fluctuations attributable to small sample sizes; (2) underlying variation in overall infant mortality rates; (3) sex-selective undercounting of children in census listings; (4) net out-migration of female children; and, finally, (5) excess female mortality in infancy and/or childhood, presumably due to discriminatory practices (Cai and Lavely 2003, 14; also Hammel, Johansson, and Ginsberg 1983; Bhaskar and Gupta 2007; cf. Courtwright 2008; Beltrán Tapia 2019). In the following section, we discuss some plausible relationships between the gender composition of the children found in the data and some of these factors. We purposely focus on components 1–3 – i.e., those that could have an impact on the variability in CSRs through random noise, different mortality environments, and data quality – which could lead to an unequal number of boys and girls even in the absence of culturally-specific gender bias in parental investment in offspring (e.g., Bender 2011).[16] The subsequent modeling section jointly assesses these factors in a multivariate framework.

Random variation

A comparative discussion of the factors that could affect CSRs must begin with random variation. Whereas child sex ratios tend to be quite homogenous at a societal level, they can be extremely noisy when computed based on small populations. Thus, such computations may result in artificially high (or low) figures simply by chance. While random fluctuations tend to narrow with increasing sample size due to the law of large numbers (Visaria 1967, 133; James 1997; Henry and Blum 1988, 15), the 95% confidence interval for the sex ratio still ranges between 99.3 and 111 even for a population as large as 10,000 individuals (and assuming a theoretical sex ratio of 105) (Guilmoto and Oliveau 2007, 5). This issue is particularly sensitive in our case, since roughly one-third of our regional populations had fewer than 1,000 children below the age of five (see section A in the Supplemental).

Instead of using point estimates, addressing this issue requires us to rely on interval measures of the sex ratios. To assess how random variation affects the corresponding CSR values in our data, we used bootstrapping techniques. Accordingly, we have re-estimated the child sex ratios and their corresponding 95% confidence intervals based on resampling with replacement.[17] Following this procedure, it is clear that while small sample sizes result in relatively wide confidence intervals, a few dozen locations still show abnormally high CSRs, even after considering the bottom part of the bootstrapped confidence interval (see Fig 1B, as well as Table 1B in the Supplemental).[18] As this approach is very conservative in a statistical sense, it should be stressed that even if the confidence interval of other extreme sex ratios is very wide, this does not necessarily mean that the observed high values are a statistical artifact. Rather, it suggests that on purely statistical grounds, we cannot rule out the possibility that the values are simply the result of random noise.

Although taking strictly statistical properties into account is critical when seeking to explain the sex ratio variation in our data, gauging their actual relevance requires comparing the bootstrapped results with appropriate region-specific reference benchmarks (the “natural” CSR in the absence of discrimination). These can be devised by considering other factors that also shaped the observed values without being directly related to the willful neglect of female offspring. While it is difficult to imagine a society that does not influence sex-specific mortality rates in one way or another (Klasen 1994, 1063), in the next step, we consider the possibility that the observed sex ratios additionally depend on the variation in the overall infant mortality and/or faulty enumerations (e.g., due to sex-selective under-registration).

Variability due to underlying infant mortality effects

While there are several demographic features that can have a bearing on the variations in child sex ratios, differential infant mortality is regarded as the quantitatively most important determinant (Coale 1991; Klasen 1994; Klasen and Wink 2003; Chao et al. 2019; cf. Hollingshaus et al. 2019). In most contemporary populations, the number of males born exceeds the number of females by approximately 105–106 to 100 (Visaria 1967; Coale 1991; Hollingshaus et al. 2019). However, a point that is often overlooked is that child sex ratios should be lower in the high-mortality environments of the past due to the female biological advantage (Klasen 1994; Klasen and Wink 2003, 269–271; Cai and Lavely 2007, 109; Beltrán Tapia and Gallego-Martínez 2017; Beltrán Tapia 2019). In circumstances in which females have access to the same nutrition and health care as males, they have greater resistance to disease throughout life, and lower mortality across all age groups (Zarulli et al. 2018).[19] However, the male vulnerability becomes more visible in high-mortality environments, at least in absolute terms: i.e., more boys die in utero, at birth, and during the first years of life, thus pushing the "natural" child sex ratio downwards. Therefore, higher mortality rates (especially during infancy) should result in lower sex ratios for infants and children.

There is indeed a clear negative link between infant mortality rates and child sex ratios. Information from 25 European countries between 1750 and 2001 shows that, as we move back in time, CSRs decrease as infant mortality rates increase (Beltrán Tapia and Gallego-Martínez 2017; Beltrán Tapia 2019). For example, while infant mortality rates of around 150 deaths per 1,000 live births correspond to a CSR of around 102 boys per 100 girls, IMRs of 220 would correspond to CSRs of around parity (100 boys per hundred girls) – and this figure could be even lower in more extreme mortality environments (see Figure 2C in the Supplemental). Therefore, it could be argued that the excessively high or low child sex ratios in our data could to some extent be a function of differential infant mortality: i.e., the ratios were high in locations where infant mortality was particularly low, and low in locations where the opposite was the case.

In order to account for this crucial dimension, we have collected a novel set of information on the overall infant mortality rates for nearly all our locations.[20] Except in a small minority of cases in which low infant mortality and high child sex ratios tend to develop in tandem, the absolute majority of our data do not conform to the expected pattern (see Figure 2C in the Supplemental). In particular, most of the extreme sex ratios are much higher than their corresponding levels of infant mortality would predict. As many of these locations can actually be characterized as high-mortality environments, their sex ratios should be much lower, and therefore cannot be explained by "natural" differential mortality between males and females (a milder mortality environment allowing more boys to survive).[21] Thus, in those locations, other factors must be at play.

The quality of the census enumeration and sex-selective under-registration

As our data are chronologically and spatially dispersed, they may be prone to significant regional variations due to the institutional arrangements surrounding the census-taking, the rationale for the enumeration, as well as the qualifications of the personnel involved in the process – all of which could affect the quality of the statistics (Thorvaldsen 2017). The accuracy of the enumeration records could vary depending not only on the individual predispositions and inclinations of the priests, estate managers, or municipal authorities responsible for maintaining them, but also on the attitudes of the respondents themselves, many of whom were illiterate, and who may at times have had various reasons not to disclose who was living with them (Szołtysek 2015, v. 2).[22] Furthermore, while there were definitely strong administrative incentives in the past to conduct a thorough registration of all individuals, including children of both sexes (e.g., Kaiser 1992, 39; Mols 1954–1956, vol. 1, 75–102), the various parties involved in preparing and drawing up particular enumerations (local governors, estate managers, clergymen) may have differed in terms of the coercive measures they had available to ensure that the statistical materials fully mirrored the populations being surveyed (e.g., Emigh, Riley, and Ahmed 2020).

Other factors likely influenced these circumstances as well. The size of the population to be enumerated must have heavily affected the final outcome, especially given the generally low organizational capacities of historical "census-takers", and a host of other challenges they faced. The chances of omissions and of undercounts (both general, as well as sex-specific) must have been minor in relatively small parishes where the priests or vicars could additionally rely on the parallel registration of vital events to double-check the information obtained from the census returns. Furthermore, the geographic location of certain populations may have affected the data collection process, especially in places where accessing some communities was difficult due to hostile biogeographic conditions, such as a rugged terrain (Diebolt and Hippe 2016; Jimenez-Ayora and Ulubaşoğlu 2015; Szołtysek, Poniat, and Gruber 2018; also Bolton 1980).

Across all these contexts, problems with census reporting may have appeared with varying intensities. For example, certain categories of individuals may have been omitted, especially children, and infants in particular. If the under-registration of these individuals occurred at random, the problems for data analysis would be substantially reduced. However, under-enumeration is often selective (Szołtysek 2015, 830). Evidence from both contemporary developing economies and historical societies suggests that adult and elderly females are more likely than males to be under-enumerated in census or census-like data (Szołtysek 2015, 890 ff; also Coleman 1976, 49; Bolton 1980, 113; Derosas 2012). Thus, if the youngest girls were not reported in these counts, high sex ratios would merely reflect their absence.[23] On the other hand, conscription or certain taxes affecting the male population may have incentivised families to hide boys rather than girls. Indeed, in many premodern censuses (especially those of a poll tax type), it was sons, not daughters, who were taxed or subject to military conscription. Therefore, it is equally possible that any under-reporting in the census would have affected males more than females (Sieff et al. 1990, 26; Szołtysek 2015, 890 ff).

Although under-registration (both general and sex-specific) during the first years of life (especially during infancy) can be a substantial concern when computing CSR values, older age groups should be less prone to this potential problem. If girls in infancy and early childhood were nominally missed by the enumerators, they should be visible in the censuses as they grew up, thus reducing sex ratios at later ages. Therefore, complementing the analysis of sex ratios at ages 0–4 with those of older age groups (i.e., 1–5 or 5–9) should help to alleviate these concerns.

Modeling historical child sex ratios

Variables and model specification

In the previous section, we discussed separately the different factors that could affect the variation in CSRs across our data. However, random variation, infant mortality, and the quality of the underlying censuses may simultaneously affect sex ratios. For example, while problems with the registration system may inflate sex ratios if girls are under-enumerated, a high-mortality environment should have the opposite effect due to the greater vulnerability of males. Therefore, the net effect would not only be ambiguous, and it might be further confounded by the presence of random noise if the sample size was not large enough. Moreover, these considerations do not preclude the possibility that gender discrimination may have also had an independent effect on sex ratios by affecting sex-specific mortality rates. Thus, it is clear that understanding the impact of the proximate factors on the observed CSR requires considering all of them simultaneously.

Therefore in what follows, we use multivariate regression to control for the impact of the different factors discussed above on the variation in CSRs across the samples. The main goal of using regression analysis is not to search for causal explanations, or to find the most powerful determinants of the observed variation. Rather, our aim is to assess whether the observed CSR in a particular region was high or low given a theoretical prediction that simultaneously takes into account the random variability, the mortality environment, and the quality of the census (earlier, Miller 1989). If we find that these factors do not fully explain the variation, we then need to consider behavioral explanations for some of the unbalanced sex ratios observed in our dataset. As we mentioned above, this is a conservative research strategy because, before attributing this variation to the outright neglect of females, it first considers other potential determinants of child sex ratios.

Accordingly, we estimate the following model:

$$\boldsymbol{CSR_i} = \boldsymbol{\alpha} + \boldsymbol{\beta}\,\boldsymbol{IMR_i} + \gamma\,\boldsymbol{QC_i} + \varepsilon_i \qquad (1)$$

where the child sex ratio in each location is regressed on the infant mortality rate and the set of variables proxying for the quality of the census discussed below. The effect of random variation is addressed by using a generalized linear model (GLM) fitted via maximum-likelihood that assumes a binomial distribution and relies on a logit function. This approach takes into account the underlying sample size (children of the corresponding age group), and therefore controls for the varying role than chance can play in determining sex ratios in different samples (Wilson and Hardy 2002).

For modeling purposes, we prefer to use the proportion of males as a dependent variable, because of its statistical properties (contrary to sex ratios *sensu stricto*, its distribution is symmetrical and follows a well-behaved distribution). Assuming that the sex of an individual is a random draw, the proportion of males (or females) follows a binomial distribution that can be approximated by a normal distribution (see Wilson and Hardy 2002; also Garenne 2008). By considering the relative number of boys and girls in different age groups (0, 0–4, 1–5, and 5–9), this exercise mitigates some of the concerns raised in the previous section regarding the potential limitations of particular sex ratios.

The choice of explanatory variables results organically from the discussion above. While it is partly guided by recent analyses of the quality of historical census data (Szołtysek, Poniat, and Gruber 2018), it also reflects the limitations of available statistical sources. On the one hand, infant mortality rates control for the greater vulnerability of males to high-mortality environments (due to data constraints, the number of observations drops slightly: $N = 308$).[24] On the other hand, a set of proxies attempts to capture the quality of the enumeration and thus potential under-registration. Relying on proxy measures is unavoidable, since without the post-enumeration checks commonly used to assess contemporary census quality, or the possibility of relying on parish registers to assess the degree of sub-registration in each census sample, it is hard to formally assess whether under-enumeration affected girls more than boys (see Griffiths, Matthews, and Hinde 2000; also Visaria 1967; Miller 1989).

The potential impact of census quality is captured by four different measures. The contextual information provided by the data inventories of our samples has been used to divide them into three groups according to the criteria of the quality of census management as suggested by the Statistical Congress of 1853 (Levi 1854, esp. 5; see Table 1B in the Supplemental for the full classification). The first group, "Modern state censuses" (29% of our regions; the reference category), identifies those counts that were carried out by "special agents, or enumerators", and for which a rule that information should be collected on a set of individual characteristics (place and date of birth, as well occupation) was clearly formulated. The second group, "Pre-modern state censuses" (44% of the regions), was also carried out by various sorts of clerical or semiclerical staff but lacked the level of detail of more modern censuses (e.g., regarding the date or place of birth). Finally, the "Other" category (27%) encompassed all of the remaining listings, particularly various types of church lists of parishes and manorial estate listings. The underlying expectation is that increased control over the management of the census (i.e., the more direct and more intensive involvement of trained personnel in the census-taking process) should greatly mitigate various types of under- and/or misreporting, including sex-selective under-enumeration. Since the classification of the census quality relies heavily on its temporality (traditional, pre-modern, modern), it partially includes the potential effects of time in our pooled cross-sectional data. In any case, we have explicitly considered the time dimension associated to our historical samples in order to further capture the possibility that the quality of the registration changed over time.

Second, our model includes the relative importance of infants and children in the population. Children were often under-registered in historical censuses and this issue may have especially affected girls and thus affect our results. Under-registration was particularly problematic for infants and this is clearly visible in some of our locations where the size of that cohort (ages 0–1) is abnormally small. As well as potentially escaping enumeration, they may have been reported as having a different age: infants who were reported as being one year old may have been registered in the 1–2 age group.[25] Therefore, including these two variables controls for the possibility that the under-registration of children was biased against girls, and may help to explain the high sex ratios. In addition, it is plausible that the census-takers faced special difficulties accessing communities located in areas with rugged terrain, which may have affected the accuracy of the counts. In order to take this dimension into account, we have considered the ruggedness index (in logs).[26]

To further control for potential enumeration issues, we need to consider the importance of age-heaping (Szołtysek, Poniat, and Gruber 2018). Although there is little evidence that there is excessive age-heaping or age displacement among the children (cf. Ewbank 1981), or that there is any correlation between age-heaping and the relative number of boys and girls in our data, the female disadvantage in age-heaping among the adult population may still serve as the litmus test for a more general gender bias in census registration. In order to account for the presence of such a bias, a female-to-male ratio in age-heaping has been computed using the Total Modified Whipple's Index (henceforth W_{tot}; see Spoorenberg 2007).[27] It is plausible to argue that a higher F/M ratio in age-heaping may thus point to more general biases in census reporting, and especially to the less precise registration of women. While at older ages such a registration bias would result in higher levels of age-heaping among women than among men, in the lower age groups (e.g., 0–4), it could lead to female-specific under-registration.

Given that urban and rural areas might differ both in terms of the disease environment and the quality of the registration, the model also controls for this dimension.[28] On the one hand, cities tended to suffer higher infant and child mortality rates and thus exhibit lower child sex ratios due to male vulnerability (Beltrán Tapia 2019; Beltrán Tapia and Gallego-Martínez 2020). On the other hand, larger locations might make the full enumeration of the population difficult, especially in overcrowded neighborhoods, thus potentially increasing under-registration. However, although less congested, it is true that rural areas may have less organizational capacities, so the net effect of the urban-rural distinction on the quality of the registration is unclear.

Finally, given that we are analyzing spatial data, it is likely that our regressions are influenced by spatial autocorrelation, which might bias both the coefficient estimates and the standard errors (Bivand, Pebesma, et al. 2013). Therefore, we have assessed whether the

Table 1. Baseline regression results.

	SR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	−0,0090	0,0194	0,6437	
Infant Mortality Rate	−0,0001	0,0000	0,0008	***
Other censuses	−0,0037	0,0201	0,8534	
Pre-modern censuses	−0,0752	0,0045	0,0000	***
Infants (over children aged 0-4)	0,1978	0,0357	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	−0,1359	0,0281	0,0000	***
Ruggedness (log)	0,0198	0,0016	0,0000	***
F/M Age-heaping (Wtot)	−0,0151	0,0071	0,0337	*
Rural	0,0823	0,0081	0,0000	***
Before 1800	−0,0248	0,0202	0,2181	
After 1850	−0,0072	0,0063	0,2585	
D2	0,528			
Moran's *I*	0,289***			
n	308			

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

model residuals are affected by this issue by computing the Moran's *I* index.[29]

Regression results

Table 1 reports the results of estimating equation 1 using CSRs (aged 0–4) as a dependent variable. Even though our model is purposely stripped down to account for IMR and under-registration issues only, it explains a substantial portion of the variation in the CSR (53 per cent).[30] As expected, the infant mortality rate and the rural dimension are significantly associated with CSRs due to the female biological advantage. In those locations where the mortality environment was harsher, the absolute gap between the number of male and female deaths was larger due to boys being more vulnerable than girls to adverse conditions (especially during infancy), which pushed the CSR downwards. The variable capturing the quality of pre-modern enumerations is also statistically significant, but its effect does not conform to our expectations since more modern censuses exhibit higher sex ratios. The female-to-male ratio in age-heaping is negatively related to the CSR, which suggests that, ceteris paribus, the adult female disadvantage in age reporting is associated with lower sex ratios. This result is also unexpected, and thus reinforces the claim that female under-registration is not a crucial issue. On the other hand, and as expected, higher terrain ruggedness is shown to be related to more masculine sex ratios. Moreover, the effects of the under-reporting of infants and children do not conform to our expectations. Other things being equal, locations with more encompassing registration of infants tend to have somewhat higher sex ratios, whereas the percentage of children aged 0–10 (relative to the working-age population) does not show statistically significant results. Finally, the variables capturing the time dimension of our data shows no association with our variable of interest. Replicating the regression using sex ratios at different age groups (0, 1–5 and 5–9) basically confirms the results reported here (see Table A1 in the appendix).

Apart from the coefficients of specific variables, the model also yields the predicted values, that is, the expected child sex ratio in each population after netting out the effect of infant mortality levels and the quality of the enumeration (and other factors potentially correlated with them, but not included in the model). This measure, depicted with a solid red line in Figure 3, may therefore be considered as a reference benchmark when assessing the "normality" of the child sex ratios observed in each location.[31] However, as mentioned earlier, random variation makes direct comparisons of CSRs from samples of different size very challenging. Therefore, instead of referring to the usual point estimates, we rely on interval estimates produced by the bootstrapping procedures already described (the whiskers in Figure 3). When the "benchmark" and the interval do not overlap, the difference between the predicted and the observed CSR can be considered statistically significant. In regions where this is the case, the relative number of boys per hundred girls is significantly higher than what would be expected based on the model. According to Figure 3, there are 48 such regions – i.e., places where the lower bound of the bootstrapped confidence interval is higher than the CSR value predicted by the regression. The absolute difference can be as large as 12 points in Russia, Hungary, and Albania (the latter would also have the largest difference observed in our data, almost 22 points), but in most cases, it is much lower (below 4.2 points). Nevertheless, these results suggest that in certain areas of historical Europe, discriminatory practices may have unduly increased female mortality rates, resulting in abnormally high CSRs.

Figure 4 charts these results geographically by depicting the model residuals of each location; i.e., the deviation between the observed CSRs and the values that the model predicts (the "relative benchmark"). The analysis of the residuals allows us to identify those locations where the CSRs are still high even after filtering out the effect of the variables included in the model. Given that the magnitude of the residuals might be affected by the sample size, we also mark those locations where the observed and the predicted values are statistically different using the procedure applied in Figure 3.[32]

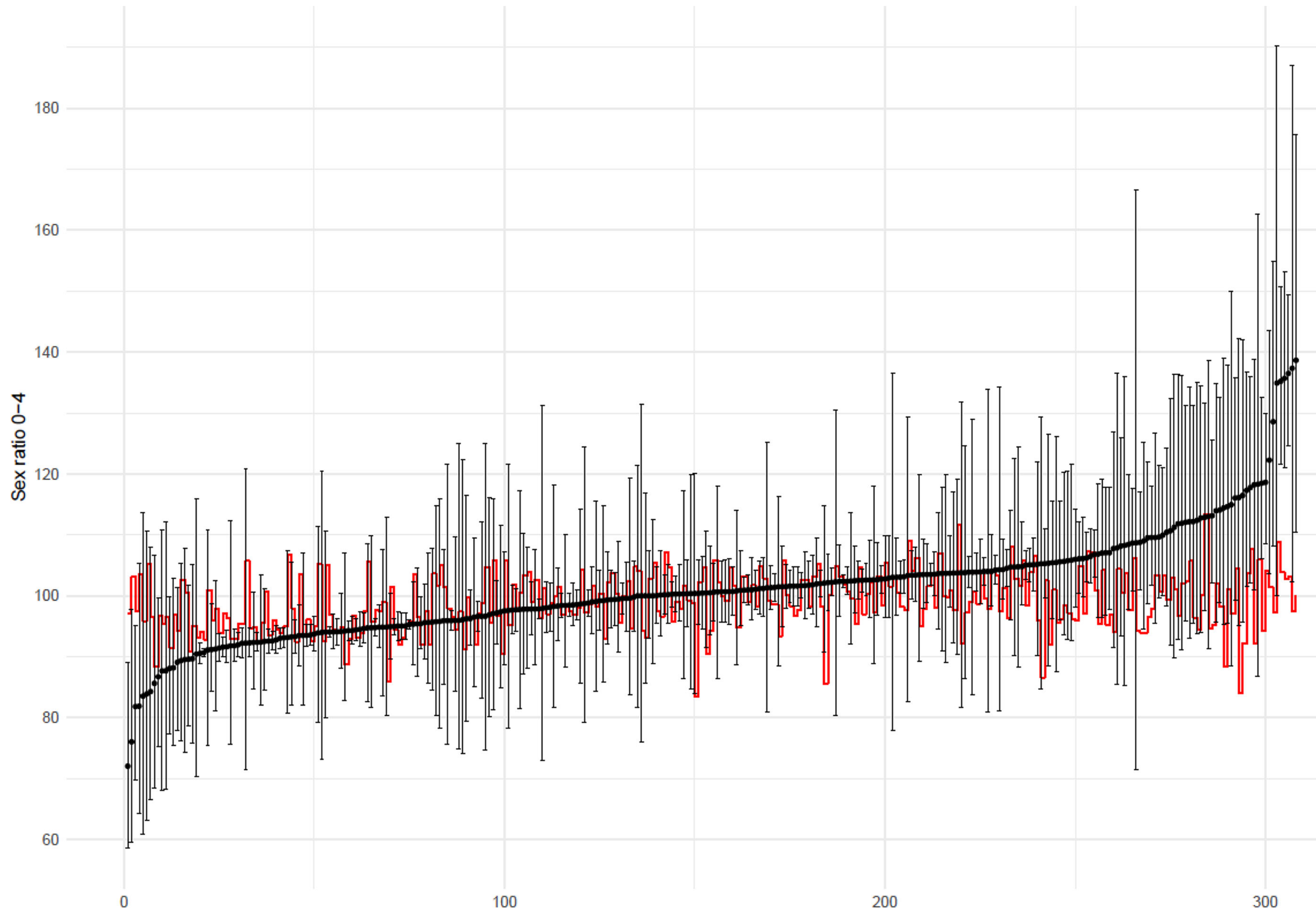

Figure 3. NAPP/Mosaic regions by bootstrapped sex ratios with 95% confidence intervals and CSRs predicted from the model (red line).
Note: the x-axis refers to the individual number identifying the locations employed in the regression model (from 1 to 308).
Source: Mosaic/NAPP data

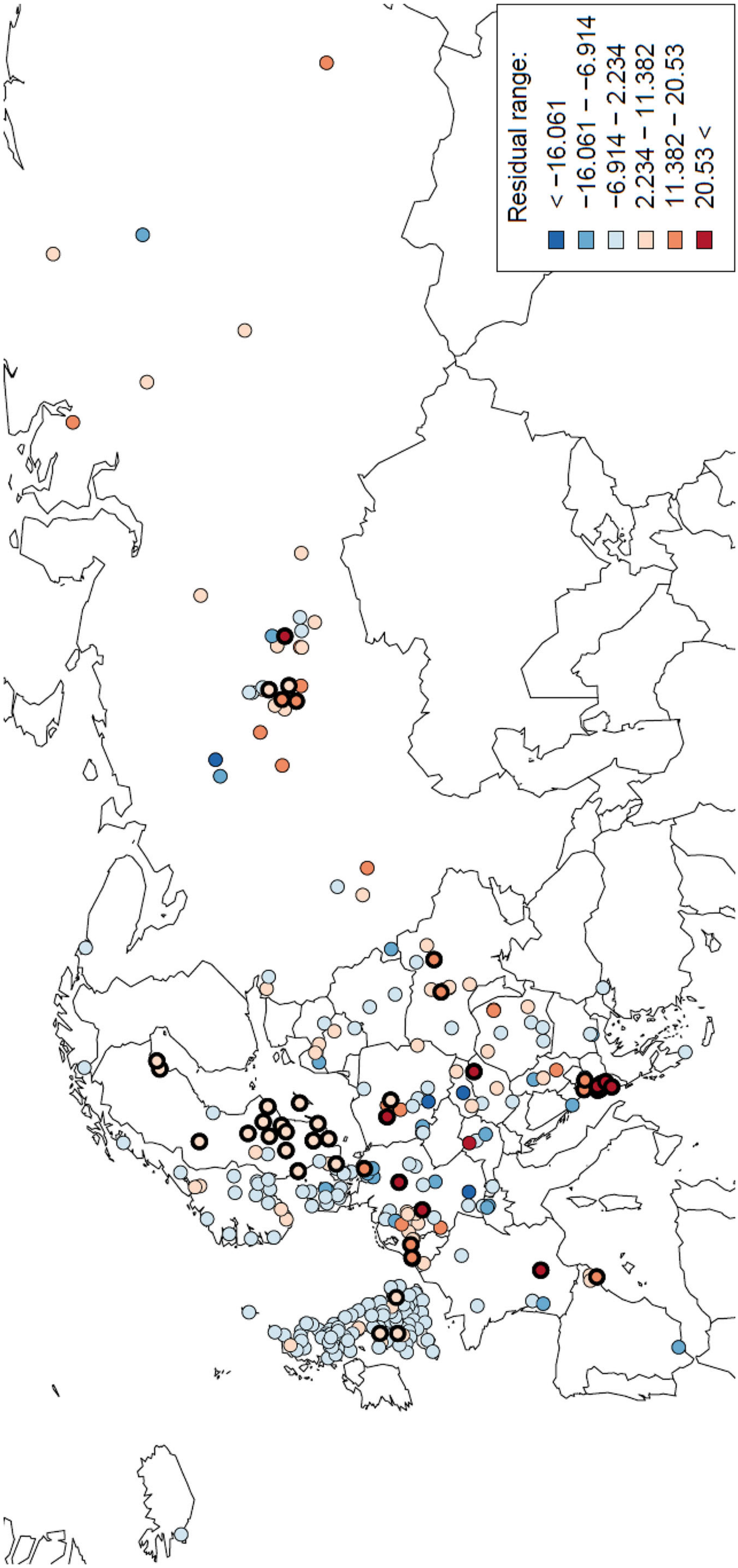

Figure 4. Spatial distribution of the model residuals.
Note: the solid circle around the point indicates regions (N = 54) where the lower bound of the bootstrapped confidence interval is higher than the CSR value predicted by the regression.
Source: Mosaic/NAPP data.

The largest departures from the values that would be predicted by the joint effect of infant mortality, random variation and census quality can be observed in those locations where residuals are above one standard deviation from the mean (dark orange and red dots in Figure 4; N = 40). These regions are geographically quite dispersed, as they are located not only to the east of our data distribution (in the Balkans, in east-central and eastern Europe, as well as in parts of Siberia),[33] but also in western Germany and the Netherlands. In most of these regions, the differences between the actual CSR values and those expected based on the model are quite large, although in some of them (in Westphalia, Upper Austria, Serbia, Romania, parts of the Ukraine, and the Urals), the differences are not statistically significant (the predicted CSRs in these regions are higher than the lower bound of the bootstrapped CIs). By contrast, most of the regions with residuals that are more than two standard deviations above the mean (red dots; N = 13) represent significant departures from the model predictions. These populations are found in the Balkans (especially in Albania) and in parts of Russia near the Urals, but also in southern France and in some scattered locations in the central German territories and Poland. For those populations, being able to explain the observed male-skewed CSRs by random variation, infant mortality, and the quality of the census seems particularly unlikely, which suggests that female neglect might also play a role.

It should be noted that 27 of the 73 populations initially put forward as exceeding the contemporary "neutral" standards (see section 2.3.) have ended up in the group for which we observe relatively large and significant departures from the model predictions. This implies that although a lion's share of the apparent regional peculiarities has been filtered out by controlling for the basic proximate factors, some populations continue to stand out as having abnormally masculinized sex ratios.

Although another group of regions in Figure 4 (N = 26) exhibit smaller (positive) residuals, and are therefore closer to the reference benchmark (less than one standard deviation from the mean; light orange), these deviations are nonetheless statistically significant. Comprising most of the Swedish populations, as well as some Norwegian, Danish, and a few English regions, their observed sex ratios are still significantly larger than the model-based benchmark (especially in Sweden). While further research is needed to explain this finding, it does open up the possibility of inquiring about the presence of gender discriminatory practices in at least some of those populations.[34]

It should also be noted that although a visual inspection of Figure 4 does not suggest that there are clear geographical patterns, the model residuals are spatially correlated (Moran's $I = 0{,}289^{***}$).[35] Our finding that the local deviations from the predicted values are correlated to those of their neighbors further suggests that behavioral factors associated with how sons and daughters were treated in these societies help to explain the variation in child sex ratios found in our sample of historical locations.

Robustness tests

The exercise provided in the previous section is potentially susceptible to several shortcomings. First, spatial autocorrelation may bias the model coefficients, as well as the residuals. Second, combining data from two sources of very different quality (NAPP vs Mosaic) may affect the results due to the larger weight of the regions located in north-western Europe (cf. Figure 1). These issues are addressed in the following section, which tests the robustness of our baseline results by: (1) including Moran's eigenvectors, and (2) excluding the NAPP locations.

Spatial autocorrelation

As evidenced by the reported Moran's coefficient, our results are influenced by spatial autocorrelation. This may affect the accuracy of the coefficient estimates, the standard errors and the model residuals (Bivand, Pebesma, et al. 2013). In order to mitigate these problems, we have re-estimated our main model using Eigenvector Spatial Filtering (Thayn 2017) (see Table A2 in the appendix). This approach removes the spatial dependence in the error of the model by choosing a set of vectors that represents the spatial autocorrelation present in the residuals and adding them to the model (Bivand, Pebesma, et al. 2013; Bivand and Piras 2015),[36] as well as effectively controlling for unobserved factors that are common across neighboring locations. However, filtering spatial autocorrelation using this procedure shows that the estimated coefficients hardly differ from our baseline specification (as expected, the explanatory power of the model increases). More importantly, the model residuals also remain largely unchanged: the correlation between the residuals from the original and the spatially filtered model is very high (r = 0,86). The similarities between the geographic distribution of the residuals from the spatially filtered model and the baseline results is further evidenced in Figure D1 (Supplemental). Note that despite including a parsimonious subset of significant spatial eigenvectors in the model, the difference between predicted CSR values and those indicated by the lower bound of the bootstrapped confidence interval continues to be statistically significant in 28 regions.

Excluding the NAPP sample

Our dataset is composed of data from the NAPP and Mosaic projects. Due to the underlying characteristics of many of its constitutive listings, the Mosaic data are more subject to quality problems.[37] Moreover, the relative weight of the NAPP information in the whole dataset, both in terms of the number of locations and the sizes of those samples, may further influence the results. In order to mitigate these concerns, we have re-estimated the main model excluding the NAPP locations. This exercise constitutes a highly challenging check, not only because of the underlying quality of the Mosaic samples, but also because it significantly reduces the variation in both the dependent and the independent variables (for instance, compared to the Mosaic populations, the NAPP locations tend to have both lower CSRs and significantly lower infant mortality rates).

The main results of this analysis are reported in Table A3 in the appendix (see also Table D2 and Figure D2 in the Supplemental). Although some of the coefficients are changed in response to the very different nature of the underlying data,[38] the model residuals remain remarkably similar to those obtained from the full model in terms of both intensity and geographic distribution (the correlation between the residuals from the full model and from the Mosaic-based model is 0,87).

The role of patriarchy

An additional source of concern is that our baseline model is likely to overestimate the potential role of the variables it includes and may therefore lead us to reach the misleading conclusion that gender discrimination was playing a negligible role in the observed sex ratios. As the estimated coefficients can be affected by other factors that are not included in the model, but that can actually trigger female excess mortality early in life, they may suffer from the well-known omitted variable bias. For example, while relatively high sex ratios observed in samples obtained from low-quality censuses might indeed be related to enumeration issues, they could also be attributable to other factors present in those societies, but that are not properly accounted for in the main specification. If that is the case, the estimated coefficient on census quality would also be partly capturing the effect of female neglect on CSRs. Likewise, apart from a healthier mortality environment, rural areas may exhibit a higher number of boys due to a more accentuated son preference that leads to gender discrimination and female excess mortality, so its coefficient might be partly capturing behavioral factors.

Therefore, a well-specified model should attempt to capture, in a distinct variable, the potential role that gender discrimination may play in shaping child sex ratios. By measuring the degree of sex- and age-related inequality across different family settings, the Patriarchy Index (Szołtysek et al. 2017; henceforth PI) provides such a proxy (see Table D4 in the Supplemental).[39] Adding this variable to our previous model has two advantages. Firstly, it allows us to better identify the specific effects of other factors that could have influenced the sex ratios. Secondly, it allows to observe, separately, the effect of female neglect on CSRs.

Table A4 in the appendix presents the results of this exercise, which indeed increases the explanatory power of our model (R-squared = 0,62). Interestingly, after netting out the effects of random noise, the mortality environment, and the potential under-registration of females, the Patriarchy Index shows a strong independent effect on the child sex ratios. Although the coefficients of most variables originally included changed only slightly, the effect of rural areas is reduced, thus evidencing that son preference is increasing female mortality in rural areas exhibiting patriarchal values. These results therefore clearly suggest that part of the variation in our dependent variable is due to gender discriminatory practices that increased excess female mortality and the sex ratios of the surviving cohorts. Replicating this exercise using the sex ratios for different age groups does not alter the results reported here. If anything, the effect of the PI is even higher for the 1–5 and 5–9 age groups, which is telling given that these groups are even less subject to potential enumeration problems (see Table D3 and Figure D3 in the Supplemental). Likewise, it should be stressed that, again, the residuals of the main model including the index are remarkably similar to our baseline model in terms of both the intensity and the geographic distribution: i.e., the correlation between the two sets of residuals is 0.93. Note that the difference between predicted CSR values and those indicated by the lower bound of the bootstrapped confidence interval continues to be statistically significant in 44 regions.

Conclusions

Inferring that high child sex ratios are indicative of excess female mortality is not straightforward, especially when the ratios are derived from historical census records. This indicator can be very random when computed from small samples, and unbalanced figures can also arise due to faulty registration and/or differences in the mortality environment. Using a novel census dataset of historic Europe, this article provides evidence that some of these regional populations exhibited high child sex ratios, often well beyond levels that are usually considered "natural". By taking a conservative approach to analyzing these observed values, our research shows that, as anticipated, part of this variation can be attributed directly to random noise associated with a small sample size (the population of children under age five from which the child sex ratios are derived). This research also finds that structural explanations related to infant mortality differentials and census quality can indeed help to explain the variation in CSRs. However, our results crucially demonstrate that in a few dozen of our locations, the observed values of CSRs appear to be too high to be solely attributable to random variation, infant mortality, or the quality of the census. These results hold regardless of the selected age group, thus mitigating potential concerns regarding sex-specific under-registration or migration, and even when the analysis is restricted to Mosaic locations.

By showing that a significant fraction of the variation in CSRs in historical Europe cannot be explained by those factors alone, we suggest that behavioral factors related to discrimination against girls likely played a role in particular regions. In this regard, the Patriarchy Index is shown to be positively associated with CSRs even after controlling for the potential influence of the set of variables mentioned above. Although our analysis is based on a restrictive set of historical locations, and our results therefore cannot be directly extrapolated to wider regions, the results indicate that the relative number of boys was abnormally high in the Balkans and in the eastern portion of European Russia, as well as in southern France and in other scattered locations in central Europe. In some other regions, and especially in Sweden, the interval measures of the CSRs are still higher than expected on the basis of our model.

These results expand on those of recent studies that suggest that gender discriminatory practices resulting in missing girls in historical Europe were more prevalent than was previously thought, especially in eastern and southern Europe. Given that the dataset analyzed here hardly touches on southern Europe, our results offer glimpses of similar behavior that may have been happening in other European regions as well. Further research is, however, needed in order to better substantiate our findings. It is important to keep in mind that our approach was purposely conservative: the model we used was very simple, and we acknowledge that it might have problems (spatial autocorrelation, omitted variables). It is also clear that adding further variables would improve the prediction (i.e., various environmental variables could be employed to proxy for various agrarian regimes). Furthermore, we may not have been able to entirely circumvent the circularity problem, as there were probably factors that affected both gender-based mortality and neglect, as well as source bias (and perhaps other explanatory factors that we did not consider). Thus, the "benchmarks" we have derived from our regression analyses would probably be more insightful if we also included other variables that could help to explain the CSR levels.

Nevertheless, both the research strategy and the results presented here underscore the need to continue discussing the most appropriate standards against which to evaluate gender mortality discrimination in historical populations. Indeed, our contribution is not just factual, it is methodological, and may therefore provide guidelines for future research. The implications of our exercise may be limited by both the quality of the underlying data and the variables included in the model. Therefore, the availability of larger samples, especially for eastern and southern Europe, and a more refined understanding of the mechanisms affecting the observed sex ratios should shed further light on the intensity of gender discrimination across historical Europe. More research is definitely needed to identify the economic, environmental, social, and cultural features that may trigger the variation in the sex ratios. For example, analyzing the individual-level information contained in historical censuses is likely to offer crucial glimpses into the types of familial or co-residential circumstances that were more associated with discrimination against girls. Given how fundamental the notion of patriarchy is to the growing body of work on the determinants of sex ratios at birth, infancy, and childhood (Basu and Das Gupta 2001), a more comprehensive exploration of the interactions of the proximate factors discussed above with various measures of family-driven age- and gender-related inequalities and environmental vicissitudes would be particularly promising. Lastly, our work cannot distinguish whether female excess mortality resulted from neglect right after birth and/or from an unequal allocation of resources during infancy and childhood. Further research should attempt to disentangle these issues, as well as providing qualitative material fleshing out the way discrimination happened.

While we agree that it would be useful to have continuous time-series data for all populations in our database, such information is unlikely to become available any time soon for a pan-European analysis such as the one carried out here. Likewise, even if available, such data would have their own limitations.[40] Still, focusing on particular case-studies relying not only on census data and/or vital statistics, but also on longitudinal information from parish registers could provide further insights into the mechanisms behind the unbalanced child sex ratios discussed in this article.

Notes

1. However, there have always been scholars who, at odds with this mainstream narrative, have argued that even in the West, some families might have chosen to keep or discard their children on the basis of their sex (see Derosas 2012 and Hanlon 2016, for a discussion).
2. For convenience, in what follows we use "sex ratios" as a general term for the number of males to females at birth, and/or in infancy and childhood, whereas we reserve the term "child sex ratios" to denote male-to-female ratios at ages 0-4 (see the Methods section). Other scholars have studied gender discrimination using heights (e.g., Baten and Murray 2000).
3. The term "natural" sex ratios refers to sex ratios in populations among whom there are no social and cultural conditions differentially affecting the survival of males and females. Sex ratios at birth are relatively stable around the level of 105 boys per 100 girls in contemporary developed countries. As this imbalance tends to decline as infants grow older due to the female biological advantage in survival, especially in high-mortality environments, the "natural" child sex ratio (aged 0-4) should be lower (Miller 1989; Beltrán Tapia and Gallego-Martínez 2017). Moreover, there is little existing research on what sex ratios at birth looked like in the past, especially considering that a higher risk of miscarriage may have affected males and females differently (Woods 2009; cf. Johansson and Nygren 1991).
4. Such a distinction reflects the discussion of the proximate determinants of elevated sex ratios in the contemporary demographic literature (e.g. Cai and Lavely 2003, 14; Cai and Lavely 2007, 108-109), with the exception that sex selective abortions were absent in pre-1950 populations. Note that elevated child sex ratios do not necessarily imply that parents were intentionally killing or neglecting their daughters. Excess female mortality at younger ages can arise from sex-differentials in parental treatment without the clear intention of getting rid of girls.
5. Please note that this paper is only focused on the androcentric bias with respect to the sex of the offspring. Other forms of imbalance are thus beyond the scope of this article.
6. There is also some evidence of male-biased sex ratios in the antebellum American west (Hammel, Johansson, and Ginsberg 1983), as well as in multiple small-scale societies (e.g., Bolton 1980; Krupnik 1985).
7. Note, however, that the norms and values prevalent in a given society may shape *both* registration practices (or biases) *as well as* gender-specific mortality neglect as such.
8. In order to enter the database, all listings (especially those from the Mosaic collection) had to pass stringent data structure evaluations (see Szołtysek and Gruber 2016).
9. In order to account for the sex proportions of all children of a certain age, we include not just family households, but also domestic units representing various kinds of institutions (poor houses, manor houses, houses of farmhands).
10. In choosing the NAPP data, we gave preference to the oldest available censuses for Iceland, Denmark, Norway (18th to early 19th centuries), and England (with Wales) (1851); whereas the earliest NAPP data for Sweden came from the late 19th century (1880). The data for Scotland came from 1881 instead of 1851, because for the latter census it was impossible to derive infant mortality estimates from around the census date. For all NAPP data, including for Great Britain, full-count populations were used. All British data came from the censuses provided to NAPP/IPUMS-I by the I-CeM project: https://icem.data-archive.ac.uk/#step1.
11. Child abandonment was widespread in historical Europe (Ransel 1988; Fuchs 1992; Kertzer 1993). Although single mothers were mostly driven by shame, the incentives to abandon boys or girls might have been different for married couples depending on their perceived value, especially in poverty-constrained contexts.
12. The range of CSRs extends from the high 130s in several regions to the low 70s in few others. Altogether, regions with CSRs below 90 – indicating a serious, if not erroneous, scarcity of boys – are very few in number (26).
13. While some of our CSRs may appear extremely high by contemporary standards, they are not empirically impossible. In fact, the highest CSR recorded in our dataset (i.e., 138.8 in north-eastern Hungary in 1869) is still far lower than the extremely masculinised child sex ratios found in contemporary China (150-197, county-level means; Cai and Lavely 2007). Values around or above 115 are also commonly observed in contemporary developing countries and even in some

European regions (Grogan 2018; also Visaria 1967), as well as among various small-scale foraging societies (Krupnik 1985).

14. Notably, the extremely masculinised regions in the east do not include the Belarusian populations, which are considered some of the most patriarchal societies of historical Europe (Szołtysek 2015).
15. Hellie's (1982) assertion that female infanticide was common in early modern Russia has been severely questioned; e.g., by Mironov (1984, 202-203; also Levin 1986).
16. Although it is plausible that a society could experience unbalanced sex ratios through migration and differential labour demand (Hammel et. 1983; Sonnino 1994) (factor 5), this type of bias does not affect our CSRs, since hardly any children <5 were autonomous migrants (Miller 1989). While it is true that in the majority of urban places in early modern Europe there are signs of at least a modest feminisation of the overall population, this feature disappears when the focus is on younger age groups (0-19) bound by the overrepresentation of males (see Fauve-Chamoux 1998).
17. For each region a random selection of individuals from the observed sample of children was drawn (the bootstrapped sample size is equal to the original, observed value). Since the procedure allows for replacement (i.e. each individual can be redrawn several times), we obtain 5,000 CSRs values for each region with the expected value reflecting the original CSR of the original sample. Then we look at the obtained distribution of CSRs to select the 2.5 and 97.5 percentiles that serve as non-parametric confidence intervals. The package *Boot* in Canty and Ripley (2020) was used (see also Davison and Hinkley 1997).
18. Altogether, 37 regions were identified for which the lower bounds of the CIs were above 100, including locations in such diverse settings as Albania, the Urals, Sweden, Scotland, France, and Catalonia, among others. It is worth mentioning that 21 out of the 73 highly masculinised regions mentioned in section 2.3 belong to this group.
19. Although this is also true in contemporary societies, the impact on the sex ratio of surviving children is negligible due to extremely low mortality levels.
20. Given the lack of harmonised, large-scale, and high-resolution evidence on the variability of IMR in the European past (cf. Kluesener et al. 2014), most of our figures are based on the available regional statistics and the voluminous secondary literature (see section C of the supplemental material). Despite the heterogenous nature of this effort, the data collected are generally consistent with spatial distribution and the evolution of infant mortality in historical Europe (see Figure 1C and Table 1C in the Supplemental; cf., e.g., Corsini and Viazzo 1997).
21. Note, however, that IMR are often considered a proxy for standard of living, so this measure can actually capture aspects of the analysed populations other than just their mortality environment.
22. While population counts were free, inclusion in birth and death registers usually involved a registration fee. Thus, there were no pecuniary incentives to hide family members in the census.
23. The possibility that female offspring were under-registered or that the redactors of our listings simply ignored many female children as essentially unimportant for the purposes of the surveys seems intuitively likely in cultures in which the role of women was seen as inferior to that of men due to strong virilocal (patrilocal/patrilineal) and exogamy norms, lineage ideology, and/or male inheritance (Bolton 1980; Bender 2011; Aldashev and Guirkinger 2012; Shi and Kennedy 2016; Szołtysek 2015). It should be noted, however, that as well as causing female under-registration, these features may have led to more direct discriminatory practices affecting the survival of girls. This further reinforces the conservative nature of our research strategy, an argument that we will follow up later.
24. Although we cannot account for the presence of unobserved mortality shocks, our measure of infant mortality would partly reflect the presence of hungers, epidemics, etc. It should be noted that, even if the presence of these shocks were biasing our econometric exercise, our results would only provide a minimun estimation of the extent of female neglect. Due to the female biological advantage, this kind of events would especially affect male children, thus pushing child sex ratios down and therefore making it more difficult to detect extremely high sex ratios.
25. Note that there are censuses in our collection that by definition did not assign any children to the <1 age group.
26. To derive information on terrain ruggedness, we used the terrain ruggedness index (Wilson et al. 2007) applying the focal function in the R library raster. Data were obtained from the GTOPO30 elevation raster dataset, which is a global digital elevation model with a horizontal grid spacing of 30 arcseconds.
27. W_{tot} considers preference for and avoidance of all 10 digits, rather than only those based on rounding one's age on a number ending with a five or a zero, while retaining the linearity and rectangularity over a five-year age range and the 23–62 age interval principles of the original Whipple's Index. W_{tot} has been shown to be most suitable for capturing digit preference in the NAPP data considered here (see Szołtysek, Poniat, and Gruber 2018).
28. In Mosaic, regions were divided by default into rural and urban categories. For the Danish and Swedish data, the official IPUMS-I rules were followed, which – as in Mosaic – identified rural populations based on the legal or administrative status of a settlement. For Norway's 1801 census, we used the variable "TownNO" and counted all people not living in one of the coded 26 towns as being rural. Iceland in 1703, given that it had no city (and no information about urban or rural residence), it was treated as wholly rural. For the British censuses considered in this research, the IPUMS-I codes were followed, which defined households "as urban if they are located in a parish with a population density of more than 75 persons per acre" (as in the Integrated Census

Microdata Project that provided the British data for NAPP; see Higgs et al. 2013, 176). In order to ensure measurement comparability, all Mosaic and relevant NAPP data with dichotomous meausres of rurality have been turned into proportions.

29. The Moran's *I* computes the correlation between the value of a particular variable y in region *i*, and the value of the (weighted) mean of *y* in neighbouring regions *j* (the five nearest neighbours were considered).
30. In generalised linear models, the square deviance (D2) is the equivalent of the more conventional R2.
31. These predicted values ("relative benchmarks") show substantial variability across our dataset, with the minimum value being as low as 84, and the maximum value being 111 boys per 100 girls. The mean predicted value is 98,53 (SD = 4,41), and is, therefore, well below the standards customarily used to gauge the "natural" level of sex ratio at birth in contemporary populations.
32. This is crucial, as the magnitude of a residual relies on the distance between the predicted and the observed values, but understood as the point estimates.
33. These results are not totally unexpected due to the recent masculinization of births in Eastern Europe (Guilmoto and Duthé 2013).
34. Child sex ratios in many Swedish locations in 1880 are higher than what our model predicts. This is perhaps unexpected and we should stress that these results are obviously contingent on the data and methodology applied here. Interestingly, however, Johansson (1984, 477-483) argues that, contrary to the previous and following period, Sweden exhibited excess female mortality (at ages 5-9 and 10-14) during the second half of the 19th century. A recent study on Scania in Sweden has suggested that, despite the female biological advantage, girls may have suffered more than boys from short-term economic stress in the 1815-1865 period (see Bengtsson 2004, 153-154). In addition, province child sex ratios (aged 0-4) in 1880 actually correlate quite well with those in 1890 which highlights the regional diversity within Sweden. This correlation between the censuses, however, disappears in 1900 thus suggesting that something was changing at the turn of the century (results available upon request). See also Karlsson et al. (2021) who argue that Sweden has not always been so gender-equal as it is today and only made substantial improvements in closing the gender gap well into the 20th century (this research also shows that there were significant differences between counties). Moreover, these regional patterns are also evident across the Swedish-Norwegian border. Swedish and Norwegian provinces along the border also share similar child sex ratios (Beltrán Tapia 2019, map 2). These spatial patterns, as well as the reasons behind them, indeed merit further attention, but fully exploring them goes beyond the scope of this paper which adopts a pan-European perspective. On the "Surplus Woman Problem" in the British census of 1851, see Levitan 2008.
35. A LISA analysis shows spatial clusters of child sex ratios that are higher than those predicted by the model in the Balkans, Scotland, Netherlands/Germany and contemporary Poland.
36. This is implemented by the ME function present in the spatialreg R package (Bivand, Pebesma, et al. 2013; Bivand and Piras 2015).
37. According to the criteria discussed above, the share of "lowest quality" listings in the Mosaic data is three times as high as the share in the NAPP data.
38. While "infants over children 0-4" had the wrong sign in the full model, this variable now conforms to our expectations. Having more infants reduces sex ratios up to age five, which would point to female under-registration. This problem seems to no longer apply to the 5-9 age group, which makes sense given that this group was less subject to under-enumeration. In this regard, the coefficients on the "type of census" also lose significance when analysing the 5-9 age group. Finally, the coefficient on the IMR also changes and becomes positive. Thus, excluding the NAPP locations with the lowest IMRs prevents an accurate assessment of the relationship between IMRs and CSRs. In addition, it is likely that the IMRs in the Mosaic data are also capturing other dimensions that have behavioural effects on the sex ratios.
39. The version of the PI used here excludes the child sex ratio component present in the original form of the index in order to avoid circularity.
40. Indeed, recording errors and distortions are common to both types of sources. parish registers also have the disadvantage that they miss the individuals who migrate out of the area of study, so they also present selection bias.

Acknowledgements

A preliminary version of this article was presented at the Hungarian Demographic Research Institute and the online seminar "Sex Ratios and Missing Girls in History". We would like to thank the participants for their comments, and Rick Mourits in particular. We also thank three anonymous reviewers and the Editor for their comments and suggestions, which were very helpful for refining the original manuscript. Miriam Hills is acknowledged for language editing. Financial support from the Research Council of Norway is also gratefully acknowledged.

Funding

This study is funded by the Research Council of Norway (project number 301527).

ORCID

Mikołaj Szołtysek http://orcid.org/0000-0003-4814-5586
Bartosz Ogórek http://orcid.org/0000-0003-3387-7936
Siegfried Gruber http://orcid.org/0000-0001-5513-4020
Francisco J. Beltrán Tapia http://orcid.org/0000-0002-7354-5699

References

Aldashev, G., and C. Guirkinger. 2012. Deadly anchor: Gender bias under Russian colonization of Kazakhstan.

Explorations in Economic History 49 (4):399–422. doi: 10.1016/j.eeh.2012.07.006.

Basu, A., and M. Das Gupta 2001. Family systems and the preferred sex of children. In N. J. Smelser & B. Baltes (eds.), *International Encyclopaedia of the Social and Behavioral Sciences*. pp. 8–5350.

Baten, J., and J. E. Murray. 2000. Heights of men and women in 19th-century Bavaria: Economic, nutritional, and disease influences. *Explorations in Economic History* 37 (4):351–69. doi: 10.1006/exeh.2000.0743.

Bechtold, B. H. 2001. Infanticide in 19th century France: A quantitative interpretation. *Review of Radical Political Economics* 33 (2):165–87. doi: 10.1177/048661340103300202.

Bechtold, B. H. 2006. The changing value of female offspring in 19th-century France: Evidence for secondary ratios. In *Killing infants: Studies in the worldwide practice of infanticide*, eds. B. H. Bechtold and D. C. Graves, 315–35. Lewiston NY: Edwin Mellen Press.

Beltrán Tapia, F. J. 2019. Sex ratios and missing girls in late-19th-century Europe. EHES Working Paper 160.

Beltrán Tapia, F. J., and D. Gallego-Martínez. 2020. What explains the missing girls in 19th-century Spain? *The Economic History Review* 73 (1):59–77. doi: 10.1111/ehr.12772.

Beltrán Tapia, F. J., and D. Gallego-Martínez. 2017. Where are the missing girls? Gender discrimination in 19th-century Spain. *Explorations in Economic History* 66:117–26. doi: 10.1016/j.eeh.2017.08.004.

Beltrán Tapia, F., and F. Marco-Gracia. 2021. Death, sex and fertility: Female infanticide in rural Spain, 1750–1950. *European Review of Economic History*. doi: 10.1093/ereh/heab023.

Beltrán Tapia, F. J., and M. Raftakis. 2021. Sex ratios and gender discrimination in Modern Greece. *Population Studies* 1–18. doi: 10.1080/00324728.2021.1923787.

Bender, T. 2011. The case of the missing girls: Sex ratios in fifteenth-century Tuscany. *Journal of Women's History* 23 (4):155–75. doi: 10.1353/jowh.2011.0049.

Bengtsson, T. 2004. Mortality and social class in four Scanian parishes, 1766–1865. In *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700–1900*, eds. T. Bengtsson, C. Campbell, J. Z. Lee, 135–71. Cambridge, MA: MIT Press.

Bhaskar, V., and B. Gupta. 2007. India's missing girls: Biology, customs, and economic development. *Oxford Review of Economic Policy* 23 (2):221–38. doi: 10.1093/oxrep/grm016.

Bivand, R. S., E. Pebesma, and V. Gomez-Rubio. 2013. *Applied spatial data analysis with R*. 2nd ed. New York: Springer. http://www.asdar-book.org/.

Bivand, R. S., J. Hauke, and T. Kossowski. 2013. Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis* 45 (2):150–79. doi: 10.1111/gean.12008.

Bivand, R. S., and G. Piras. 2015. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software* 63 (18):1–36. doi: 10.18637/jss.v063.i18.

Bolton, R. 1980. High-altitude sex ratios: How high? *Medical Anthropology* 4 (1):107–43. doi: 10.1080/01459740.1980.9965865.

Cai, Y., and W. Lavely. 2003. China's Missing Girls: Numerical Estimates and Effects on Population Growth. *The China Review* 3 (2):13–29.

Cai, Y., and W. Lavely. 2007. Child sex ratios and their regional variation in China. In *Transition and challenge: China's population at the beginning of the 21st century*, eds. Z. Zhongwei and F. Guo, 108–23. London: OUP.

Canty, R. A., and B. Ripley. 2020. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-25.

Chao, F., P. Gerland, A. R. Cook, and L. Alkema. 2019. Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proceedings of the National Academy of Sciences of the United States of America* 116 (19):9303–11. doi: 10.1073/pnas.1812593116.

Coleman, E. 1976. Infanticide in the early middle ages. In *Women in medieval society*, ed. by S. M. Stuard, 47–70. Philadelphia: University of Pennsylvania Press.

Charpentier, A., and E. Gallic. 2020. La démographie historique peut-elle tirer profit des données collaboratives des sites de généalogie? *Population* 75 (2):391–421. doi: 10.3917/popu.2002.0391.

Coale, A. 1991. Excess female mortality and the balance of the sexes. *Population and Development Review* 17 (3):517–23. doi: 10.2307/1971953.

Corsini, C. A., and P. P. Viazzo, eds. 1997. *The decline of infant and child mortality: The European experience, 1750–1990*. The Hague: Martinus Nijhoff Publishers.

Courtwright, D. T. 2008. Gender imbalances in history: Causes, consequences and social adjustment. *Reproductive BioMedicine Online* 16:32–40. doi: 10.1016/S1472-6483(10)60397-5.

Das Gupta, M., J. Zhenghua, L. Bohua, Z. Zhenming, X. Chung, and B. Hwa-Ok. 2003. Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *The Journal of Development Studies* 40 (2):153–87. doi: 10.1080/00220380412331293807.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

den Boer, A. M., and V. M. Hudson. 2006. Sex-selective infanticide and the "Missing Females" in China and India. In *Killing Infants: Studies in the worldwide practice of infanticide*, eds. B. H. Bechtold, and D. C. Graves, 337–71. Lewiston NY: Edwin Mellen Press.

Delille, G. 1974. Un problème de démographie historique: Hommes et femmes face à la mort. *Mélanges de L'Ecole Française de Rome. Moyen-Age, Temps Modernes* 86 (2):419–43. doi: 10.3406/mefr.1974.2316.

Derosas, R. 2012. Suspicious deaths: Household composition, infant neglect, and child care in nineteenth-century Venice. *Annales de Démographie Historique* 123 (1):95–126. doi: 10.3917/adh.123.0095.

Derosas, R., and N. O. Tsuya. 2010. Child control as a reproductive strategy. In *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900*, ed. N.O. Tsuya, 129–55. Cambridge: Cambridge University Press.

Diebolt, C., and R. Hippe. 2016. Remoteness equals backwardness? Human capital and market access in the European regions: Insights from the long run. Working Papers of BETA 2016-32, Bureau d'Economie Théorique et Appliquée, UDS, Strasbourg.

Emigh, R. J., D. Riley, and P. Ahmed. 2016. *Antecedents of censuses from medieval to nation states. How societies and states count*. New York, Palgrave Macmillan.

Emigh, R., D. Riley, and P. Ahmed. 2020. The sociology of official information gathering: Enumeration, influence, reactivity, and power of states and societies. In *The new handbook of political sociology*, eds. T. Janoski, C. De Leon, J. Misra, and I. William Martin, 290–320. Cambridge: Cambridge University Press.

Ewbank, D. C. 1981. *Age misreporting and age-selective underenumeration: Sources, patterns, and consequences for demographic analysis*. Washington, DC: National Academy Press.

Fauve-Chamoux, A. 1998. Le surplus urbain des femmes en France préindustrielle et le rôle de la domesticité. *Population (French Edition)* 53 (1–2):359–77. doi: 10.2307/1534249.

Fellman, J. 2015. Glimpses at the history of sex ratio studies. *Science Journal of Public Health* 3 (2):291–302. doi: 10.11648/j.sjph.20150302.30.

Fossett, M. A., and K. J. Kiecolt. 1991. A methodological review of the sex ratio: Alternatives for comparative research. *Journal of Marriage and the Family* 53 (4):941–57. doi: 10.2307/352999.

Fuchs, R. G. 1992. *Poor and pregnant in Paris: Strategies for survival in the nineteenth century*. New Brunswick: Rutgers University Press.

Garenne, M. 2008. Heterogeneity in the sex ratio at birth in European populations. *Genus* 64 (3–4):99–108.

Gini, C. 1908. *Il sesso dal punto di vista statistica: Le leggi della produzione dei sessi*. Bologna: Remo Sandron Editore.

Greenhalgh, S. 2013. Patriarchal demographics? China's sex ratio reconsidered. *Population and Development Review* 38:130–49. doi: 10.1111/j.1728-4457.2013.00556.x.

Griffiths, P., Z. Matthews, and A. Hinde. 2000. Understanding the sex ratio in India: A simulation approach. *Demography* 37 (4):477–88. doi: 10.1353/dem.2000.0004.

Grogan, L. A. 2018. Strategic fertility behaviour, early childhood human capital investments and gender roles in Albania. https://ssrn.com/abstract=2615504.

Guilmoto, C. Z. 2018. Sex ratio imbalances in Asia: An ongoing conversation between anthropologists and demographers. In *Scarce women and surplus men in China and India*, eds. S. Srinivasan, and S. Li, 145–61. Cham: Springer.

Guilmoto, C. Z., and G. Duthé. 2013. Masculinization of births in Eastern Europe. *Population & Societies* 506 (11):1–4.

Guilmoto, C. Z., and S. Oliveau. 2007. Sex ratio imbalances among children at micro-level: China and India compared. Paper presented at Population Association of America 2007 Annual Meeting, March, New York. <halshs-00296636>

Hammel, E. A., S. R. Johansson, and C. A. Ginsberg. 1983. The value of children during industrialization: Sex ratios in childhood in nineteenth-century America. *Journal of Family History* 8 (4):346–66. doi: 10.1177/036319908300800404.

Hanlon, G. 2003. L'infanticidio di coppie sposate in Toscana nella prima età moderna. *Quaderni Storici* 38:453–98.

Hanlon, G. 2016. Routine infanticide in the West, 1500–1800. *History Compass* 14 (11):535–48. doi: 10.1111/hic3.12361.

Hanlon, G. 2017. Death control in the West: New research on routine infanticide in Northern Italy from the 16th to the 18th century. (unpublished manuscript; version: July 24, 2017).

Henry, L. 1968. The verification of data in historical demography. *Population Studies* 22 (1):61–81. doi: 10.1080/00324728.1968.10405526.

Henry, L., and A. Blum. 1988. *Techniques d'analyse en Démographie Historique*. Paris: INED.

Herlihy, D., and C. Klapisch-Zuber. 1985. *Tuscans and their families: A study of the Florentine Catasto of 1427*. New Haven: Yale University Press.

Higgs, E., C. Jones, K. Schürer, and A. Wilkinson. 2013. *Integrated Census Microdata (I-CeM) Guide*. Available: https://www.nappdata.org/napp/resources/enum_materials_pdf/sample_design_gb1851a.pdf (Accessed: 22 June 2020).

Hynes, L. 2011. Routine infanticide by married couples? An assessment of baptismal records from 17th century Parma. *Journal of Early Modern History* 15 (6):507–30. doi: 10.1163/157006511X600828.

Hollingshaus, M., R. Utz, R. Schacht, and K. R. Smith. 2019. Sex ratios and life tables: Historical demography of the age at which women outnumber men in seven countries, 1850–2016. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52 (4):244–53. doi: 10.1080/01615440.2019.1605863.

James, W. H. 1997. The validity of inferences of sex-selective infanticide, abortion and neglect from unusual reported sex ratios at birth. *European Journal of Population = Revue Europeenne de Demographie* 13 (2):213–7.

Jayachandran, S. 2015. The roots of gender inequality in developing countries. *Annual Review of Economics* 7 (1):63–88. doi: 10.1146/annurev-economics-080614-115404.

Jimenez-Ayora, P., and M. A. Ulubaşoğlu. 2015. What underlies weak states? The role of terrain ruggedness. *European Journal of Political Economy* 39:167–83. doi: 10.1016/j.ejpoleco.2015.04.014.

Johansson, S., and O. Nygren. 1991. The missing girls of China: A new demographic account. *Population and Development Review* 17 (1):35–51. doi: 10.2307/1972351.

Johansson, S. R. 1984. Deferred infanticide: Excess female mortality during childhood. In *Infanticide: Comparative and evolutionary perspectives*, eds. G. Hausfater and S. B. Hrdy, 463–85. New York: Aldine.

Kaiser, D. H. 1992. Urban household composition in early modern Russia. *Journal of Interdisciplinary History* 23 (1):39–71. doi: 10.2307/205481.

Karlsson, T., J. Kok, and F. Perrin. 2021. The historical gender gap index: A longitudinal and spatial assessment of Sweden, 1870–1990. Lund Papers in Economic History 217.

Kemkes, A. 2006. Secondary sex ratio variation during stressful times: The impact of the French Revolutionary Wars on a German parish (1787–1802). *American*

Journal of Human Biology: The Official Journal of the Human Biology Council 18 (6):806–21. doi: 10.1002/ajhb.20562.

Kertzer, D. I. 1993. *Sacrificed for honor: Italian infant abandonment and the politics of reproductive control*. Boston: Beacon Press.

Klasen, S. 1994. Missing women reconsidered. *World Development* 22 (7):1061–71. doi: 10.1016/0305-750X(94)90148-1.

Klasen, S., and C. Wink. 2003. Missing women: Revisiting the debate. *Feminist Economics* 9 (2–3):263–99. doi: 10.1080/1354570022000077999.

Kluesener, S., I. Devos, P. Ekamper, I. N. Gregory, S. Gruber, J. Martí-Henneberg, F. van Poppel, L. Espinha da Silveira, and A. Solli. 2014. Spatial inequalities in infant survival at an early stage of the longevity revolution: A pan-European view across 5000+ regions and localities in 1910. *Demographic Research* 30 (68):1849–64.

Krupnik, I. 1985. The male-female ratio in certain traditional populations of the Siberian Arctic. *Études/Inuit/Studies* 9 (1):115–40.

Levi, L. 1854. Resume of the statistical congress, held at Brussels, September 11th, 1853, for the purpose of introducing unity in the statistical documents of all countries. *Journal of the Statistical Society of London* 17 (1):1–14. doi: 10.2307/2338350.

Levin, E. 1986. Infanticide in pre-petrine Russia. *Jahrbücher Für Geschichte Osteuropas (Neue Folge)* 34:215–24.

Levitan, K. 2008. Redundancy, the 'surplus woman' problem, and the British census, 1851–1861. *Women's History Review* 17 (3):359–76. doi: 10.1080/09612020801924449

Lynch, K. A. 2011. Why weren't (many) European women "missing"? *The History of the Family* 16 (3):250–66. doi: 10.1016/j.hisfam.2011.02.001.

Manfredini, M., M. Breschi, and A. Fornasin. 2016. Son preference in a sharecropping society: Gender composition of children and reproduction in a pre transitional Italian community. *Population* 4 (4):641–58.

Marco-Gracia, F. J., and F. J. Beltrán Tapia. 2021. Son preference, gender discrimination and missing girls in rural Spain, 1750–1950. *Population and Development Review* 47 (3):665–87. doi: 10.1111/padr.12406.

Miller, B. D. 1989. Changing patterns of juvenile sex ratios in rural India: 1961 to 1971. *Economic and Political Weekly* 24 (22):1229–36.

Miller, B. D. 2001. Female-selective abortion in Asia: Patterns, policies, and debates. *American Anthropologist* 103 (4):1083–95. doi: 10.1525/aa.2001.103.4.1083.

Mironov, B. N. 1984. Istorija cholopstva v Rossii v obvescenii amerikanskogo istorika. *Istorija SSSR* 3:194–206.

Mols, R. 1954–1956. *Introduction à la démographie historique des villes d'Europe du XIVe au XVIIIe siècle*. Louvain: Bibliothèque de l'Université/Bureaux du recueil.

Ransel, D. L. 1988. *Mothers of misery: Child abandonment in Russia*. Princeton: Princeton University Press.

Reynolds, G. 1979. Infant mortality and sex ratios at baptism as shown by reconstruction of Willingham, a parish at the edge of the Fens in Cambridgeshire. *Local Population Studies* 22:31–7.

Ring, R. R. 1979. Early medieval peasant households in central Italy. *Journal of Family History* 4 (1):2–21. doi: 10.1177/036319907900400101.

Ruggles, S., E. Roberts, S. Sarkar, and M. Sobek. 2011. The North Atlantic population project: Progress and prospects. *Historical Methods* 44 (1):1–6. doi: 10.1080/01615440.2010.515377.

Sandström, G., and L. Vikström. 2015. Sex preference for children in German villages during the fertility transition. *Population Studies* 69 (1):57–71. doi: 10.1080/00324728.2014.994667.

Scalone, F., and R. Rettaroli. 2015. Exploring the variations of the sex ratio at birth from an historical perspective. *Statistica* 75 (2):213–26.

Sen, A. 1990. More than 100 million women are missing. *New York Review of Books* 37:61–6.

Shi, Y., and J. Kennedy. 2016. Delayed registration and identifying the "missing girls" in China. *The China Quarterly* 228:1018–38. doi: 10.1017/S0305741016001132.

Sieff, D. F., L. Betzig, L. Cronk, A. G. Fix, M. Flinn, L. Sattenspiel, K. Gibson, D. A. Herring, N. Howell, S. R. Johansson, et al. 1990. Explaining biased sex ratios in human populations: A critique of recent studies [and comments and reply]. *Current Anthropology* 31 (1):25–48. doi: 10.1086/203801.

Sonnino, E. 1994. In the male city: The 'Status Animarum' of Rome in the seventeenth century. In *Socio-economic consequences of sex-ratios in historical perspective, 1500–1900*, eds. Antoinette Fauve-Chamoux and Solvo Sogner, 19–30. Milan: Università Bocconi.

Spoorenberg, T. 2007. Quality of age reporting: Extension and application of the modified Whipple's index. *Population (English Edition)* 62 (4):729–42. doi: 10.3917/pope.704.0729.

Szołtysek, M. 2015. *Rethinking East-central Europe: Family systems and co-residence in the Polish-Lithuanian Commonwealth (2 vols)*. Bern: Peter Lang.

Szołtysek, M., and S. Gruber. 2016. Mosaic: Recovering surviving census records and reconstructing the familial history of Europe. *The History of the Family* 21 (1):38–60. doi: 10.1080/1081602X.2015.1006655.

Szołtysek, M., and B. Ogórek. 2020. How many household formation systems were there in historic Europe? A view across 256 regions using partitioning clustering methods. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53 (1):53–76. doi: 10.1080/01615440.2019.1656591.

Szołtysek, M., B. Ogórek, R. Poniat, and S. Gruber. 2020. Making a place for space: A demographic spatial perspective on living arrangements among the elderly in historical Europe. *European Journal of Population = Revue Europeenne de Demographie* 36 (1):85–117. doi: 10.1007/s10680-019-09520-5.

Szołtysek, M., S. Klüsener, R. Poniat, and S. Gruber. 2017. The patriarchy index: A new measure of gender and generational inequalities in the past. *Cross-Cultural Research* 51 (3):228–62. doi: 10.1177/1069397117697666.

Szołtysek, M., and R. Poniat. 2018. Historical family systems and contemporary developmental outcomes: What is to be gained from the historical census microdata revolution? *The History of the Family* 23 (3):466–92. doi: 10.1080/1081602X.2018.1477686.

Szołtysek, M., R. Poniat, and S. Gruber. 2018. Age heaping patterns in Mosaic data. *Historical Methods: A Journal of*

Quantitative and Interdisciplinary History 51 (1):13–38. doi: 10.1080/01615440.2017.1393359.

Thayn, J. B. 2017. Eigenvector spatial filtering and spatial autoregression. In *Encyclopedia of GIS [electronic resource]*, eds. Shashi Shekhar, Hui Xiong, and Xun Zhou, 511–22. New York: Springer.

Thorvaldsen, G. 2017. *Censuses and census takers: A global history* (1st ed.). London: Routledge.

Visaria, P. 1967. Sex ratio at birth in territories with a relatively complete registration. *Eugenics Quarterly* 14 (2): 132–42. doi: 10.1080/19485565.1967.9987713.

Willigan, J. D., and K. A. Lynch. 1982. *Sources and methods of historical demography*. New York: Academic Press.

Wilson, M. F. J., B. O'Connell, C. Brown, J. C. Guinan, and A. J. Grehan. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy* 30 (1-2):3–35. doi: 10.1080/01490410701295962.

Wilson, K., and I. Hardy. 2002. Statistical analysis of sex ratios: An introduction. In *Sex ratios: Concepts and research methods*, ed. I. Hardy, 48–92. Cambridge: Cambridge University Press.

Woods, R. 2009. *Death before birth: Fetal health and mortality in historical perspective*. Oxford: Oxford University Press.

Zarulli, V., J. A. Barthold Jones, A. Oksuzyan, R. Lindahl-Jacobsen, K. Christensen, and J. W. Vaupel. 2018. Women live longer than men even during severe famines and epidemics. *Proceedings of the National Academy of Sciences of the United States of America* 115 (4): E832–E840. doi: 10.1073/pnas.1701535115.

Appendix

Table A1. Regression results (baseline specification, different age-groups).

	SR 0			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	−0,0185	0,0508	0,7166	
Infant Mortality Rate	−0,0001	0,0001	0,4984	
Other censuses	−0,0704	0,0668	0,2918	
Pre-modern censuses	−0,0871	0,0138	0,0000	***
Infants (over children aged 0-4)	0,3786	0,1296	0,0035	**
Children aged 0-10 (over pop. aged 15-64)	−0,1736	0,0759	0,0221	*
Ruggedness (log)	0,0006	0,0003	0,0181	*
F/M Age-heaping (Wtot)	−0,0213	0,0173	0,2189	
Rural	0,1070	0,0200	0,0000	***
Before 1800	0,0374	0,0667	0,5753	
After 1850	0,0201	0,0178	0,2570	
D2		0,2747		
Moran's *I*		0,0488		
n		280		

	SR 1-5			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	0,0025	0,0188	0,8955	
Infant Mortality Rate	−0,0001	0,0000	0,0073	**
Other censuses	−0,0076	0,0190	0,6879	
Pre-modern censuses	−0,0788	0,0043	0,0000	***
Infants (over children aged 0-4)	0,1964	0,0337	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	−0,1317	0,0272	0,0000	***
Ruggedness (log)	0,0009	0,0001	0,0000	***
F/M Age-heaping (Wtot)	−0,0124	0,0071	0,0783	•
Rural	0,0959	0,0079	0,0000	***
Before 1800	−0,0188	0,0191	0,3250	
After 1850	−0,0036	0,0070	0,6117	
D2		0,4904		
Moran's *I*		0,2612 ***		
n		308		

	SR 5-9			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	−0,0563	0,0203	0,0054	**
Infant Mortality Rate	−0,0001	0,0000	0,0051	**
Other censuses	−0,0217	0,0189	0,2503	
Pre-modern censuses	−0,0629	0,0046	0,0000	***
Infants (over children aged 0-4)	0,1328	0,0351	0,0002	***
Children aged 0-10 (over pop. aged 15-64)	−0,0267	0,0289	0,3548	
Ruggedness (log)	0,0188	0,0017	0,0000	***
F/M Age-heaping (Wtot)	−0,0058	0,0075	0,4376	
Rural	0,0863	0,0085	0,0000	***
Before 1800	−0,0004	0,0189	0,9823	
After 1850	−0,0207	0,0066	0,0018	**
D2		0,4897		
Moran's *I*		0,2812 ***		
n		308		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$

Table A2. Spatial model with Moran's eigenvectors.

	SR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	0,0050	0,0210	0,8116	
Infant Mortality Rate	−0,0001	0,0000	0,0505	•
Other censuses	−0,0265	0,0202	0,1905	
Pre-modern censuses	−0,0841	0,0046	0,0000	***
Infants (over children aged 0-4)	0,1429	0,0371	0,0001	***
Children aged 0-10 (over pop. aged 15-64)	−0,0040	0,0304	0,8966	
Ruggedness (log)	0,0124	0,0018	0,0000	***
F/M Age-heaping (Wtot)	−0,0017	0,0081	0,8346	
Rural	0,0460	0,0086	0,0000	***
Before 1800	−0,0203	0,0203	0,3164	
After 1850	−0,0420	0,0068	0,0000	***
Moran's Eigenvectors				
vec2	−0,4552	0,0621	0,0000	***
vec18	0,2570	0,0171	0,0000	***
vec33	−0,1910	0,0164	0,0000	***
vec42	0,3078	0,0752	0,0000	***
vec15	−0,1818	0,0231	0,0000	***
vec182	−0,1536	0,0219	0,0000	***
D2		0,744		
Moran's *I*		0,036		
n		308		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$

Table A3. Regression results (baseline specification, Mosaic only).

	SR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	−0,0297	0,0436	0,4965	
Infant Mortality Rate	0,0000	0,0001	0,7154	
Other censuses	−0,0707	0,0308	0,0218	*
Pre-modern censuses	−0,0847	0,0258	0,0010	**
Infants (over children aged 0-4)	−0,0757	0,0744	0,3090	
Children aged 0-10 (over pop. aged 15-64)	0,0566	0,0557	0,3097	
Ruggedness (log)	0,0003	0,0002	0,0923	•
F/M Age-heaping (Wtot)	0,0562	0,0181	0,0019	**
Rural	0,0017	0,0166	0,9193	
Before 1800	−0,0031	0,0280	0,9124	
After 1850	0,0015	0,0258	0,9549	
D2		0,2993		
Moran's *I*		0,0823		
n		158		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$

Table A4. Regression results including the Patriarchy Index (baseline dataset).

	SR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	−0,0392	0,0195	0,0446	*
Infant Mortality Rate	−0,0003	0,0000	0,0000	***
Other censuses	−0,0480	0,0203	0,0184	*
Pre-modern censuses	−0,0914	0,0047	0,0000	***
Infants (over children aged 0-4)	0,2846	0,0363	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	−0,1527	0,0281	0,0000	***
Ruggedness (log)	0,0171	0,0016	0,0000	***
F/M Age-heaping (Wtot)	−0,0532	0,0077	0,0000	***
Rural	0,0656	0,0082	0,0000	***
Before 1800	0,0004	0,0202	0,9834	
After 1850	0,0123	0,0065	0,0584	•
Patriarchy Index	0,0080	0,0006	0,0000	***
D2		0,608		
Moran's *I*		0,1750 ***		
n		308		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$