DISCUSSION PAPER SERIES

DP15818

Inferring “missing girls” from child sex ratios in European historical census data: A conservative approach

Francisco Beltrán Tapia, Mikołaj Szołtysek, Bartosz Ogórek and Siegfried Gruber

ECONOMIC HISTORY

CEPR

ISSN 0265-8003

Inferring “missing girls” from child sex ratios in European historical census data: A conservative approach

Francisco Beltrán Tapia, Mikołaj Szołtysek, Bartosz Ogórek and Siegfried Gruber

Discussion Paper DP15818
Published 16 February 2021
Submitted 16 February 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programmes:

- Economic History

Inferring “missing girls” from child sex ratios in European historical census data: A conservative approach

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 1950; 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.

JEL Classification: I14, I15, J13, J16, N33

Keywords: Sex ratios, Infant and child mortality, Gender Discrimination, health

Francisco Beltrán Tapia - francisco.beltran.tapia@ntnu.no
Norwegian University of Science and Technology and CEPR

Mikołaj Szołtysek - mszoltis@gmail.com
Hungarian Demographic Research Institute

Bartosz Ogórek - ogorek_bartosz@o2.pl
Pedagogical University of Cracow

Siegfried Gruber - si.gruber@uni-graz.at
Karl-Franzens-Universität Graz

Acknowledgements
A preliminary version of this article was presented at the Hungarian Demographic Institute and the online seminar "Sex Ratios and Missing Girls in History". We would like to thank the participants for their comments, as well as Rick Mourits for his detailed feedback. Financial support from the Norwegian Research Council (Project 90569400) is also gratefully acknowledged.

Inferring “missing girls” from child sex ratios in European historical census data: a conservative approach[1]

Francisco J. Beltrán Tapia
(Norwegian University of Science and Technology)

Mikołaj Szołtysek
(Hungarian Demographic Research Institute)

Bartosz Ogórek
(Pedagogical University of Cracow)

Siegfried Gruber
(Karl-Franzens-Universität Graz)

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 1950; 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, Infant and child mortality, Gender discrimination, Health

JEL Codes: I14, I15, J13, J16, N33

---

[1] A preliminary version of this article was presented at the Hungarian Demographic Institute and the online seminar "Sex Ratios and Missing Girls in History". We would like to thank the participants for their comments, as well as Rick Mourits for his detailed feedback. Financial support from the Norwegian Research Council (Project 90569400) is also gratefully acknowledged.

1. INTRODUCTION AND BACKGROUND

Thirty years after Amartya Sen forced the world to pay attention to the phenomenon of missing girls in South and East Asia (e.g., Sen 1990; Das Gupta et al. 2003; Jayachandran 2015; Guilmoto 2018), the question of whether similar forms of gender discrimination in historical Europe 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)[2].

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[3]. The usual method employed to deduce discriminatory practices leading to excess female mortality in infancy and childhood compares the actual age-specific sex ratios with the “expected” ones; that is, the ratios that would result from equal treatment of the sexes in the distribution of survival-related ends (Klassen 1994; Klassen and Wink 2003). However, calls for the establishment of a non-discriminatory standard against which particular historical conditions can be evaluated have been speculative at best, as we simply do not know what sex ratios in the past should have looked like. Although various “natural” sex ratio standards have been proposed for contemporary human populations (Visaria 1967; Coale 1991; Klassen 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 large-scale comparative projects[4]. This is because the combined effects of random variation, overall mortality

---

[2] 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).

[3] 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).

[4] 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

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 behavioural explanations[5].

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 variations due to demographic and environmental determinants (similarly Henry and Blum 1988; Dellile 1974; also Fellman 2015). Discussing skewed sex ratios in the Florentino census of 1427, Herlihy and Klapish-Zuber (1985, 131-135) wondered whether they were indicative of gender differentials in mortality arising from negligence or indifference towards 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 vulnerable to underlying variation in the overall infant mortality rates (Beltran-Tapia and Gallego-Martínez 2017). 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 et al. 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 the 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"

---

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 2018). 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).

[5] 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.

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 paper, 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 wilful 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?[6]

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 2020; also Manfredini et al. 2016; Sandström and Vikström 2015; Johannson 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)[7]. 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 characterised 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).

---

[6] 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.

[7] There is also some evidence of male-biased sex ratios in the antebellum American west (Hammel et al. 1983), as well as in multiple small-scale societies (e.g., Bolton 1980; Krupnik et al. 2019).

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 (Fosset 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 modelling 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 2019; Beltrán Tapia and Gallego-Martínez 2017). 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 modelling 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[8]. 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 paper 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 accepted too eagerly 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 paper 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 intensity 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 section five, we show that our baseline results hold up to a battery of robustness tests. In the final section, we further discuss our results.

---

[8] 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.

2. MATERIALS AND METHODS

2.1 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, harmonised 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, their organisation, and the types of information they provide[9]. 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[10]. 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[11]. Altogether, we collected information on nearly four million

---

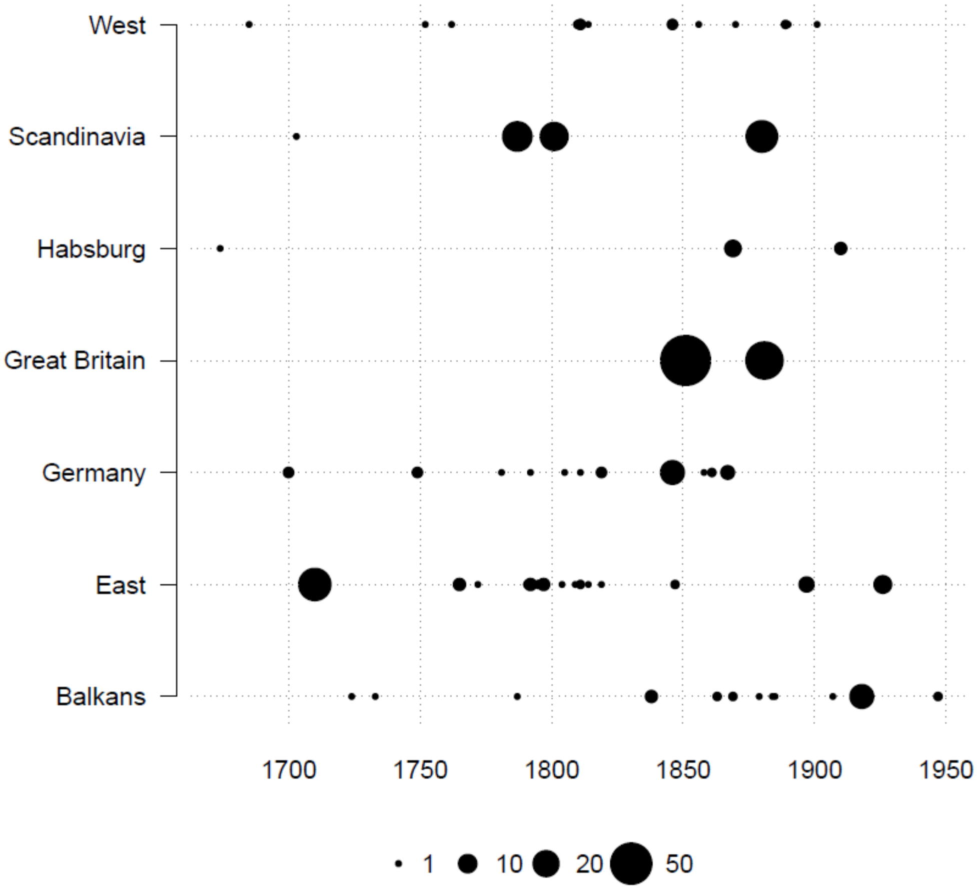
[9] In order to enter the database, all listings (especially those from the Mosaic collection) had to pass stringent data structure evaluations (see Szołtysek and Gruber 2016).

[10] 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).

[11] 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

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: 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 (online).

---

data came from the censuses provided to NAPP/IPUMS-I by the I-CeM project: https://icem.data-archive.ac.uk/#step1.

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 digitised census microdata. Section A of the supplemental material provides further evidence that these data do not compromise the analyses presented below.

2.2 The measure of sex proportions

While much of the literature on sex ratios deals with sex ratios at birth (a flow measure), this paper 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. First, 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 (Miller 1989; Agnihotri 1996; Cai and Lavely 2007; Bashkar and Gupta 2007; den Boer and Hudson 2006). Second, 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. Third, 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.

2.3 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 show evidence of unusually high sex ratios (see also Table 1B in the supplemental material)[13]. Although these regions do not form any specific cluster, they tend to congregate somewhat towards 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.

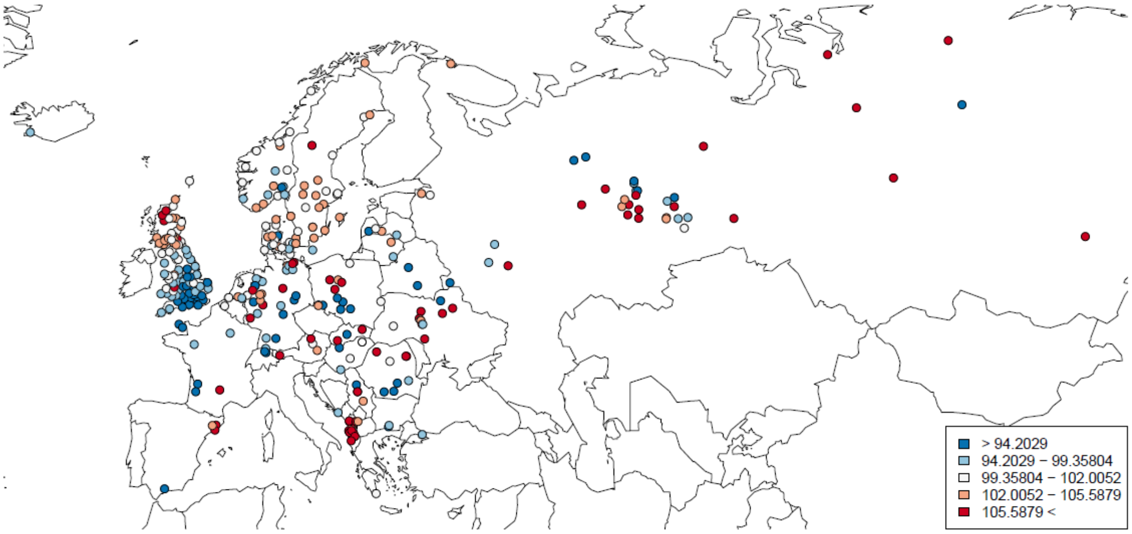
---

[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 et al. 2019).

[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).

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.

Overall, the picture remains fairly patchy, since areas with high sex ratios are often contiguous with or relatively close to areas where the CSRs are more balanced or even much 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). 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. To sum up, although the spatial distribution of the CSRs does not seem to be entirely random (the global Moran's *I* index of spatial autocorrelation is 0.27***), it remains sufficiently patchy that we can preclude the possibility that the distribution has any direct connections to well-known demographic cultures or regions. The question that has yet

---

[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).

to be answered is to what extent this fragmented picture can actually be linked to local factors.

3. 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 et al. 1983; Bashkar 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 modelling section jointly assesses these factors in a multivariate framework.

3.1 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

---

[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).

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 material).

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 material)[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 artefact. 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, to gauge their actual importance, we have to compare 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 wilful 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

---

[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.

additionally depend on the variation in the overall infant mortality and/or faulty enumerations (e.g., due to sex-selective under-registration).

3.2 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. Hollingshau 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 died 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 material). 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

---

[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.

infant mortality was particularly low, and low in locations where the opposite was the case.

In order to check whether this was the case, 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 material). 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 characterised 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.

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

As our data are chronological 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. 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–

---

[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 material; 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.

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., Emight et al. 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 organisational 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 rugged terrain (Diebolt and Hippe 2016; Jimenez-Ayora and Ulubaşoğlu 2015; Szołtysek et al. 2018; also Bolton 1980).

Across all of 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 1974, 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

---

[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.

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.

4. MODELLING HISTORICAL CHILD SEX RATIOS

4.1 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, 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 in order to understand the impact of the proximate factors on the observed CSR, all of them should be considered simultaneously.

Accordingly, 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 behavioural explanations for some of the unbalanced sex ratios observed in our dataset. As we mentioned above, this is a conservative research strategy, because before it attributes this variation to the outright neglect of females, it first considers other potential determinants of child sex ratios.

Accordingly, we estimate the following model:

$$\boldsymbol{CSR_i = \alpha + \beta\, IMR_i + \beta\, QC_i + \varepsilon_i} \tag{1}$$

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 modelling 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 in section 3. While it is partly guided by recent analyses of the quality of historical census data (Szołtysek et al. 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). 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 et al. 2000; also Visaria 1967; Miller 1989).

The potential impact of census quality is captured through 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 material 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. As this results in the redundancy of the explicit time variable in the models, this variable has therefore been omitted.

Second, our model includes the relative importance of infants and children in the population. Children were often under-registered in historical censuses. This issue is clearly visible in some of our samples, and may have affected girls in particular. Under-registration was particularly problematic for infants, since 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 not have been registered in the 0-1 age group)[24]. 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[25].

To further control for potential enumeration issues, we need to consider the importance of age-heaping (Szołtysek et al. 2018). Although there is little evidence that there is excessive age-heaping or age displacement among 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)[26]. 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.

Finally, given that we are analysing 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 et al. 2013). Therefore, we have assessed whether the model residuals are affected by this issue by computing the Moran's *I* index[27].

---

[24] Note that there are censuses in our collection that by definition did not assign any children to the <1 age group.

[25] 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.

[26] 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 et al. 2018).

[27] 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).

4.2 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 (48 per cent)[28]. As expected, the infant mortality rate is negatively 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. Finally, 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. 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, what the child sex ratio in each population is expected to look like 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[29]. However, as mentioned earlier, random variation

---

[28] In generalised linear models, the square deviance (D2) is the equivalent of the more conventional R2.

[29] 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.

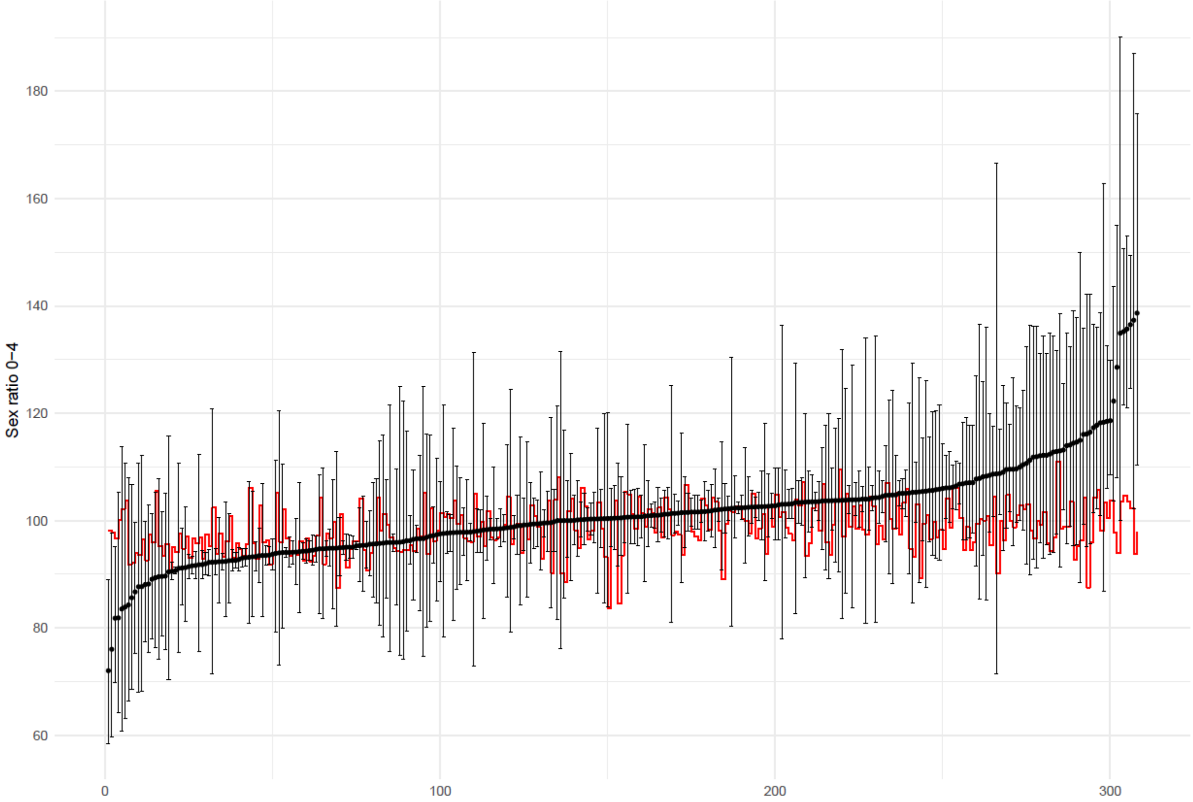
makes direct comparisons of CSRs from samples of different size very challenging. Therefore, instead of referring to the usual point estimates, we rely on interval estimates produced by the bootstrapping procedures already described (the whiskers in Figure 3). When the “benchmark” and the interval do 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 54 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, over 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.

Table 1: Baseline regression results

	SR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	0,0061	0,0164	0,7088	
Infant Mortality Rate	-0,0003	0,0000	0,0000	***
Other censuses	0,0013	0,007	0,8555	
Pre-modern censuses	-0,0683	0,0044	0,0000	***
Infants (over children aged 0-4)	0,1465	0,0354	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	0,0024	0,0245	0,9211	
Ruggedness	0,0258	0,0014	0,0000	***
F/M Age-heaping (Wtot)	-0,0219	0,007	0,0017	**
D2	0,479			
Moran's I	0.234 ***			
n	308			

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Figure 3: NAPP/Mosaic regions by bootstrapped sex ratios with 95% confidence intervals and CSRs predicted from the model (red line)



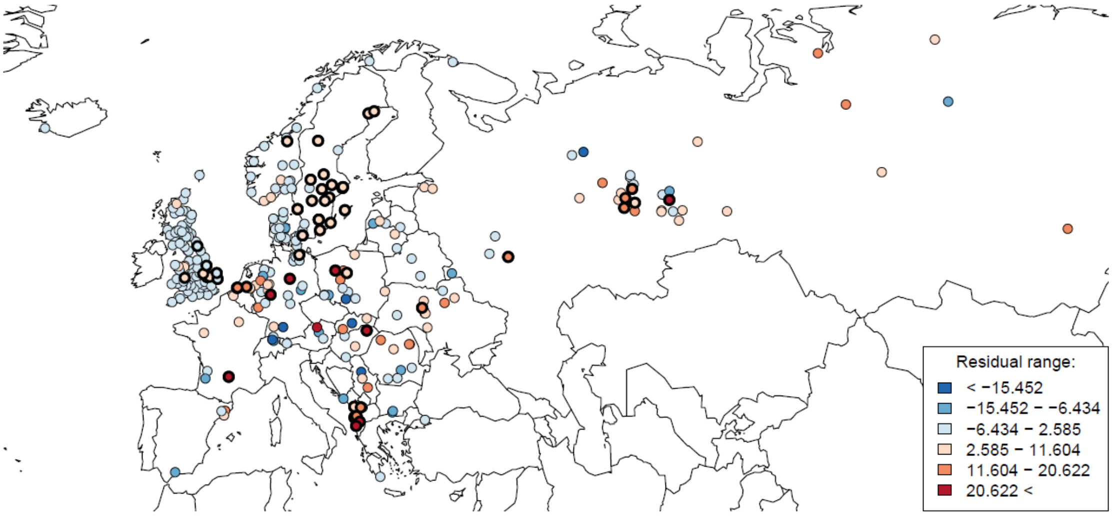
Source: Mosaic/NAPP data.

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 above[30].

---

[30] 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.

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=43). 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), 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=12) 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 only 18 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 have been filtered out by controlling for the basic proximate factors, some populations continue to stand out as having abnormally masculinised sex ratios.

Although another group of regions in Figure 4 (N=29) exhibit smaller (positive) residuals, and are therefore closer to the reference benchmark (one sd from the mean and less; 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[31].

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 = 0.234***). Our finding that the local deviations from the predicted values are correlated to those of their neighbours further suggests that behavioural 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.

5. 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). Finally, as the variables included in the model may be related to other factors that actually lead more directly to female neglect, the effects

---

[31] 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. On the “Surplus Woman Problem” in the British census of 1851, see Levitan 2008

of non-behavioural factors are likely to be overestimated. These issues are addressed in the following section, which tests the robustness of our baseline results by: (1) including Moran's eigenvectors, (2) excluding the NAPP locations, and (3) adding the Patriarchy Index to the regression.

5.1 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 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 et al. 2013; Bivand and Piras 2015)[32], as well as effectively controlling for unobserved factors that are common across neighbouring 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=0.88). 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 materials). 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 27 regions.

5.2 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

---

[32] This is implemented by the ME function present in the spatialreg R package (Bivand et al. 2013; Bivand and Piras 2015).

subject to quality problems[33]. 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 while 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 material). Although some of the coefficients are changed in response to the very different nature of the underlying data[34], 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 0.88).

5.2 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

---

[33] 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.

[34] 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.

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.

Therefore, a well-specified model should take into account 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 material)[35]. Thus, adding this variable to our previous model allows us to better identify the specific effects of other factors that could have influenced the sex ratios.

Table A4 in the appendix presents the results of such an analysis. The coefficients of the variables originally included changed only slightly. Interestingly, after netting out the effects of random noise, the mortality environment, and the potential under-registration of females, the PI shows a strong independent effect on the child sex ratios. Thus, these results 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 material). 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 0.98. 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 45 regions.

6. 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.

---

[35] 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.

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 analysing 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; i.e., 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, 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 behavioural 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 that led to missing girls in historical Europe were more prevalent than was previously thought, especially in eastern and south-eastern Europe. Given that the dataset analysed here hardly touches on southern Europe, our results offer glimpses of similar behaviour that may have been happening in other European regions as well. However, it is clear that further research is 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, analysing 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.

REFERENCES

Aldashev, G., Guirkinger, C. 2012. Deadly anchor: Gender bias under Russian colonization of Kazakhstan. Explorations in Economic History 49(4): 399-422.

Baten, J., Murray, J.E. 2000. Heights of Men and Women in 19th-Century Bavaria: Economic, Nutritional, and Disease Influences. Explorations in Economic History 37(4): 351-369.

Bechtold, B. H. 2001. Infanticide in 19th century France: A quantitative interpretation. Review of Radical Political Economics 33(2): 165–187.

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*, ed. B.H. Bechtold and D.C. Graves, 315-335. 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 Gallego-Martínez, D. 2020. What explains the missing girls in 19th-century Spain? Economic History Review 73(1): 59–77.

Beltrán Tapia, F. J., Gallego-Martínez, D. 2017. Where are the missing girls? Gender discrimination in 19th-century Spain. Explorations in Economic History 66: 117-126.

Beltrán Tapia, Francisco & Marco-Gracia, Francisco. 2020. Death, sex and fertility: Female infanticide in rural Spain, 1750-1950. EHES Working Paper | No. 186, June 2020.

Beltrán Tapia, F.J., and Raftakis, M. 2019. 'All little girls, the bad luck!' Sex ratios and gender discrimination in 19th-century Greece, EHES Working Paper 172.

Bender, T. 2011. The case of the missing girls: sex ratios in fifteenth-century Tuscany. Journal of Women's History 23(4):155-75.

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, ed. T. Bengtsson, C Campbell, J.Z. Lee et al., 135-171, MIT Press.

Bhaskar, V., Bishnupriya, G. 2007. India's missing girls: biology, customs, and economic development. Oxford Review of Economic Policy 23(2): 221–238.

Bivand, R.S., Pebesma, E., Gomez-Rubio, V. 2013. Applied spatial data analysis with R (2nd edition). Springer, NY. http://www.asdar-book.org/

Bivand, R.S., Hauke, J., Kossowski, T. 2013. Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. Geographical Analysis 45(2): 150-179.

Bivand, R.S., Piras, G. 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software 63(18): 1-36.

Bivand, R. S., Hauke, J., and Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. Geographical Analysis 45(2): 150-179.

Bolton, R. 1980. High-Altitude Sex Ratios: How high? Medical Anthropology: Cross-Cultural Studies in Health and Illness 4(1): 107-143.

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

Cai, Y., Lavely, W. 2007. Child sex ratios and their regional variation in China. In *Transition and challenge: China's population at the beginning of the 21st century*, ed. Z. Zhongwei and F. Guo, 108–123. London: OUP.

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

Chao, F., Gerland, P., Cook, A. R., Alkema, L. 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 116 (19): 9303–9311.

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., Gallic, E. 2020. La démographie historique peut-elle tirer profit des données collaboratives des sites de généalogie? Population 2(2): 391-421.

Coale, A. 1991. Excess Female Mortality and the Balance of the Sexes. Population and Development Review 17(3): 517–523.

Corsini, C.A., Viazzo, P.P. (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.

Courtwright, D. T. 1990. The neglect of female children and childhood sex ratios in the nineteenth century: a review of the evidence. Journal of Family History 25: 313–323.

Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, Z., Chung, X, and Hwa-Ok, B. 2003. Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. Journal of Development Studies 40(2): 153–187.

Davison, A. C., Hinkley, D. V. 1997. *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

De Hansen, E. G. R. 1979. "Overlaying" in 19th-century England: Infant mortality or infanticide? Human Ecology 7(4): 333–352.

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

Delille, G. 1974. Un problème de démographie historique: hommes et femmes face à la mort. Mélanges de l' École Française de Rome; Moyen Age – Temps modernes 86(2): 419-443.

Derosas, R. 2012. Suspicious deaths: household composition, infant neglect, and child care in nineteenth-century Venice. Annales de démographie historique 123(1): 95-126.

Derosas, R., Tsuya, N.O. 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 et al., 129–155. Cambridge: Cambridge University Press.

Diebolt, C., Hippe, R. 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., Riley, D., Ahmed, P. 2016. *Antecedents of Censuses from Medieval to Nation States. How Societies and States Count*. New york, Palgrave Macmillan.

Emigh, R., Riley, D., & Ahmed, P. 2020. The Sociology of Official Information Gathering: Enumeration, Influence, Reactivity, and Power of States and Societies. In T. Janoski, C. De Leon, J. Misra, & I. William Martin (Eds.), *The New Handbook of Political Sociology* (pp. 290-320). Cambridge: Cambridge University Press.

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

Fauve-Chamoux Antoinette. 1998. Le surplus urbain des femmes en France préindustrielle et le rôle de la domesticité. Population 53(1-2): 359-377.

Fellman, J. 2015. Glimpses at the History of Sex Ratio Studies. Science Journal of Public Health 3(2): 291-302.

Fossett, M.A., Kiecolt, K. J. 1991. A Methodological Review of the Sex Ratio: Alternatives for Comparative Research. Journal of Marriage and Family 53(4): 941-957.

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: 130-149.

Griffiths, P., Matthews, Z. & Hinde, A. 2000. Understanding the sex ratio in India: A simulation approach. Demography 37: 477–488.

Grogan, L. A. 2018. Strategic Fertility Behaviour, Early Childhood Human Capital Investments and Gender Roles in Albania. Available at SSRN: https://ssrn.com/abstract=2615504.

Gruber, S., Szołtysek, M. 2016. The patriarchy index: a comparative study of power relations across historical Europe. The History of the Family, 21(2): 133-174.

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*, ed. S. Srinivasan, S. Li, 145-161. Cham: Springer.

Guilmoto, C.Z., Oliveau, S. 2007. Sex Ratio Imbalances Among Children At Micro-Level: China And India Compared. Population Association of America 2007 Annual Meeting, Mar 2007, New York, United States. <halshs-00296636>

Hammel, EA, Johansson SR, Ginsberg CA. 1983. The Value of Children During Industrialization: Sex Ratios in Childhood in Nineteenth-Century America. Journal of Family History 8(4):346-366.

Hanlon, G. 2003. L'infanticidio di coppie sposate in Toscana nella prima età moderna, Quaderni Storici 38: 453–498;

Hanlon, G. 2016. Routine infanticide in the West, 1500-1800. *History Compass* 14(11): 535-548.

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.

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

Herlihy, D. 1975. Life Expectancies for Women in Medieval Society. In *The Role of Women in the Middle Ages*, ed. R.T. Morewedge, 633-56. Albany: SUNY Press.

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

Hynes, L. 2011. Routine infanticide by married couples? An assessment of baptismal records from 17th century Parma. Journal of Early Modern History 15: 507-530.

Hollingshaus, M., Utz, R., Schacht, R., Smith, K. R. 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): 244-253.

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 13(2):213-7.

Jayachandran, S. 2015. The Roots of Gender Inequality in Developing Countries. Annual Review of Economics 7(1): 63-88

Jimenez-Ayora, P., Ulubaşoğlu, M. A. 2015. What underlies weak states? The role of terrain ruggedness. European Journal of Political Economy 39: 167-183.

Johansson, S., Nygren, O. 1991. The Missing Girls of China: A New Demographic Account. Population and Development Review 17(1): 35-51.

Johansson, S.R. 1984. Deferred infanticide: Excess female mortality during childhood. In Infanticide: Comparative and Evolutionary Perspectives, ed. G. Hausfater and S.B. Hrdy, 463-485. New York: Aldine.

Kaiser, D. H. 1992. Urban Household Composition in Early Modern Russia. Journal of Interdisciplinary History, 23(1): 39–71.

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 18(6): 806–821.

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

Klasen, S., Wink, C. 2003. "Missing women": revisiting the debate. Feminist Economics 9(2-3): 263-299.

Kluesener, S. and multiple authors. 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): 1849-1864.

Krupnik, I. 1985. The Male-Female Ratio in Certain Traditional Populations of the Siberian Arctic. Études/Inuit/Studies 9(1): 115-140.

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: 215–224.

Levitan, K. 2008. Redundancy, the 'Surplus Woman' Problem, and the British Census, 1851–1861. Women's History Review 17(3): 359–376.

Lynch, K.A. 2011. Why weren't (many) European women "missing"? History of the Family 16: 250-266.

Manfredini, M., Breschi, M. & Fornasin, A. 2016. Son Preference in a Sharecropping Society: Gender Composition of Children and Reproduction in a Pre-Transitional Italian Community. Population 4(4): 641-658.

Miller, B.D. 1989. Changing Patterns of Juvenile Sex Ratios in Rural India: 1961 to 1971. Economic and Political Weekly 24 (22): 1229-1236.

Miller, B. D. 2001. Female-selective abortion in Asia: Patterns, policies, and debates. American Anthropologist 103(4): 1083–1095.

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.

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: 31-37.

Ring, R. R. 1979. Early Medieval Peasant Households in Central Italy. Journal of Family History 4(1): 2–21.

Ruggles, S., Roberts, E., Sarkar, S., & Sobek, M. 2011. The North Atlantic Population Project: Progress and prospects. Historical Methods 44: 1-6.

Sandström, G., Vikström, L. 2015. Sex preference for children in German villages during the fertility transition. Population Studies 69(1): 57-71.

Scalone, F., Rettaroli, R. 2015. Exploring the Variations of the Sex Ratio at Birth from an Historical Perspective. Statistica 75(2): 213-226.

Sen, A. 1990. More than 100 million women are missing. New York Rev. of Books 37: 61-66.

Shi, Y., Kennedy, J. 2016. Delayed Registration and Identifying the "Missing Girls" in China. The China Quarterly 228: 1018-1038.

Sieff, Daniela F., 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.

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-E 62 (4):729–742.

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

Szołtysek, M., Gruber, S. 2016. Mosaic: recovering surviving census records and reconstructing the familial history of Europe. The History of the Family 21(1): 38-60.

Szołtysek, M., Ogórek, B. 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.

Szołtysek, M., Ogórek, B., Poniat, R. *et al.* 2020. Making a Place for Space: A Demographic Spatial Perspective on Living Arrangements Among the Elderly in Historical Europe. *European Journal of Population* 36: 85–117.

Szołtysek, M., Klüsener, S., Poniat, R., Gruber, S. 2017. The Patriarchy Index: A New Measure of Gender and Generational Inequalities in the Past. Cross-Cultural Research 51(3):228-262.

Szołtysek, M., Poniat, R. 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): 466-492.

Szołtysek, Poniat, R., Gruber, S. 2018. Age heaping patterns in Mosaic data. Historical Methods: A Journal of Quantitative and Interdisciplinary History 51(1): 13-38.

Thayn, J. B. 2017. Eigenvector Spatial Filtering and Spatial Autoregression. In Encyclopedia of GIS [electronic resource], ed. Shashi Shekhar, Hui Xiong, Xun Zhou, *511–522.* Springer.

Thorvaldsen, G. 2017. *Censuses and Census Takers: A Global History* (1st ed.). Routledge.

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

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

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

Wilson, K., Hardy, I. 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., Barthold Jones, J.A., Oksuzyan, A., Lindahl-Jacobsen, R., Christensen, K., Vaupel, J.W. 2018. Women live longer than men even during severe famines and epidemics. Proc Natl Acad Sci 115(4):E832-E840.

APPENDIX

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

	CSR 0			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,0048	0,0438	0,9127	
Infant Mortality Rate	-0,0003	0,0001	0,0001	***
Other censuses	-0,0234	0,0197	0,2351	
Pre-modern censuses	-0,0774	0,0138	0,0000	***
Infants (over children aged 0-4)	0,3391	0,1292	0,0087	**
Children aged 0-10 (over pop. aged 15-64)	0,0446	0,063	0,4789	
Ruggedness	0,02	0,0033	0,0000	***
F/M Age-heaping (Wtot)	-0,0326	0,017	0,0546	.
D2	0,247			
Moran's I	0.045			
n	280			
	CSR 1-5			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,01045	0,01599	0,51321	
Infant Mortality Rate	-0,00025	0,00003	0,0000	***
Other censuses	0,01083	0,00683	0,11322	
Pre-modern censuses	-0,06473	0,00425	0,0000	***
Infants (over children aged 0-4)	0,1243	0,03334	0,00019	***
Children aged 0-10 (over pop. aged 15-64)	0,02374	0,02413	0,32536	
Ruggedness	0,0257	0,00143	0,0000	***
F/M Age-heaping (Wtot)	-0,0196	6,89E-03	0,00444	**
D2	0,462			
Moran's I	0.229 ***			
n	308			
	CSR 5-9			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,0524	0,0169	0,002	**
Infant Mortality Rate	-0,0002	0,0000	0,0000	***
Other censuses	0,0206	0,0071	0,0039	**
Pre-modern censuses	-0,0545	0,0045	0,0000	***
Infants (over children aged 0-4)	0,0724	0,0346	0,0365	*
Children aged 0-10 (over pop. aged 15-64)	0,1049	0,0256	0,0000	***
Ruggedness	0,0262	0,0015	0,0000	***
F/M Age-heaping (Wtot)	-0,0135	0,0073	0,0000	.
D2	0,43			
Moran's I	0.255 ***			
n	308			

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

Table A2: Spatial Model with Moran's Eigenvectors

	CSR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	0,0250	0,0191	0,1893	
Infant Mortality Rate	-0,0001	0,0000	0,0000	***
Other censuses	-0,0210	0,0072	0,0037	**
Pre-modern censuses	-0,0605	0,0051	0,0000	***
Infants (over children aged 0-4)	0,0513	0,0369	0,1643	
Children aged 0-10 (over pop. aged 15-64)	0,0391	0,0260	0,1336	
Ruggedness	0,0075	0,0018	0,0000	***
F/M Age-heaping (Wtot)	-0,0089	0,0079	0,2597	
Moran's Eigenvectors				
vec18	0,1186	0,0179	0,0000	***
vec2	-0,3653	0,06033	0,0000	***
vec7	0,2530	0,0492	0,0000	***
vec33	-0,1537	0,0160	0,0000	***
vec42	0,2771	0,0751	0,0002	***
vec4	-0,2306	0,0343	0,0000	***
vec8	0,1511	0,0735	0,0399	*
D2	0,745			
Moran's I	0,047			
n	308			

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

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

	CSR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,0180	0,0360	0,6161	
Infant Mortality Rate	0,0000	0,0001	0,8213	
Other censuses	-0,0794	0,0169	0,0000	***
Pre-modern censuses	-0,0916	0,0191	0,0000	***
Infants (over children aged 0-4)	-0,0651	0,0754	0,3875	
Children aged 0-10 (over pop. aged 15-64)	0,0422	0,0529	0,4251	
Ruggedness	0,0055	0,0052	0,2885	
F/M Age-heaping (Wtot)	0,0553	0,0170	0,0011	**
D2	0,298			
Moran's I	0.082 *			
n	160			

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

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

	CSR 0-4			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	0,0004	0,0164	0,979	
Infant Mortality Rate	-0,0004	0,0000	0,0000	***
Other censuses	-0,0462	0,0077	0,0000	***
Pre-modern censuses	-0,0894	0,0046	0,0000	***
Infants (over children aged 0-4)	0,2545	0,0361	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	-0,047	0,0247	0,0573	.
Ruggedness	0,0198	0,0015	0,0000	***
F/M Age-heaping (Wtot)	-0,0642	0,0075	0,0000	***
Patriarchy Index	0,0085	0,0006	0,0000	***
D2	0,579			
Moran's I	0.136 ***			
N	308			

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

SUPPLEMENTAL MATERIAL

A. THE NAPP/MOSAIC CENSUS MICRODATA AND THEIR POTENTIAL LIMITATIONS

The age-sex statistics as obtained from NAPP/Mosaic Data are fully harmonized across our samples and guarantee that comparable sex ratio indicators can be derived from them. Census takers asked for individual ages (not birth dates) in the overwhelming majority of the NAPP and Mosaic listings, and such ages were commonly recorded as age at the last birthday[1]. In cases where the birth dates were provided (e.g. Hungarian censuses 1869, Sweden 1880, Austria 1910) adjustments were made during the harmonization process, so that age reported refers to the age at last birthday. Unlike the majority of the censuses included in our database, the Danish (1787) and Norwegian (1801) listings (altogether 40 regions) stood apart from the rests of the collection in that the originally recorded ages were expressed as the next, but not the achieved age-year, i.e. newborns were enumerated with age 1, and people in their 26th year were written as 26 years of age and not 25[2]. However, all these ages have been recalculated by the NAPP data providers, so they refer to the age of last birthday (the originally-recorded ages are available in the unharmonized source variables stored at IPUMS-I website)[3]. As a result of the recalculation done by the NAPP, the finally obtained 0-4 child cohorts for these two censuses are oversized when compared to the original unharmonized data. However, a comparison of the CSRs (0-4) from harmonized and unharmonized data has shown that these are nearly identical.

There is obviously strong age heaping in some Mosaic and NAPP data. Although its extent varies from region to region and census to census, it mostly affects the adult population (Szołtysek et al. 2018). Children are generally free from excessive age heaping and age displacement (under-reporting, which is a separate issue, is discussed in the main text). Based on analysis of more than 150 age-sex distributions from censuses and surveys (many from developing countries), Ewbank (1981) established that the ages of infants and children tend to be reported more accurately than the age of adults, which he attributes to their reporting by parents or other adults who remember the birth and to the rapid physiological and psychological changes during childhood that makes it easier to guess the age with reasonable accuracy (p. 48 ff), and there is no reason to believe that similar circumstances had not mattered in the past. Some parents, however, may have exaggerated the ages of their children through rounding off, rather than truncating it to the number of completed years, thus leading to an overstatement by a year for those within 6 months of the reported age. As long as such mild forms of age displacement would only concern *Children* under 1 being reported as 1 year of age, they will not disturb the cumulative size of the focal 0-4 cohort, but the problem could be more serious if children with completed fourth birthday would be reported as five years old, and thus moved forward in to the old age group (Demeny and Shorter 1968; also Miller 1989). However, there is no suggestion that such a tendency was dominant across our data, as in the overwhelming majority of our regional populations four years-old cohorts were bigger than those at the age of 5. Finally, the fact that the results of the regression analyses do not change when we

---

[1] Note, however, that even if the census was requesting birth date, individuals often derived the year of birth by using their age as a reference and counting backwards (Buławski 1930).

[2] See: https://international.ipums.org/international-action/variables/AGE#questionnaire_text_section

[3] https://international.ipums.org/international-action/variables/AGE#comparability_section

employ different age-groupings (see the main text and the Appendix) further reassures that the age-heaping is not a serious issue in our case.

Other normative differences between some of the censuses (e.g., lodgers in the British and the Swedish censuses may have occasionally been treated as separate units by enumerators; unlike the majority of our data, the Danish, Icelandic and Norwegian listings followed *de iure* criteria of enumerations), have no impact on the sex ratio statistics as defined in this paper.

Two limitations of our dataset should, however, be mentioned here. First, combining full-count censuses (or samples of thereof) from NAPP with local/regional listings from Mosaic yields a considerable variation in the size of children population across the samples. This results in one third of our regional populations having less than 1,000 children below 5, a problem that specially affects locations in the Mosaic dataset. While this could imply that some of the regional variations in the sex ratios to be observed may be subject to high levels of random noise, the methodology applied in this paper has been set to explicitly account for whether the observed sex ratios fall within what would be expected according to the underlying population of children of a defined age.

The second source of complication is the unequal distribution of data across regions and time periods (see Fig. 1 in the main text), which stems from the availability of digitised census microdata. However unfortunate this mixing of time periods might be, it does not seem to introduce particular biases to our analyses. Although it may be expected that the social and economic pressure for gender discrimination should diminish over time, in the absence of manipulation or the interference of historical events affecting either males or females in a population, the child sex ratios actually tend to be remarkably stable in human populations across long stretches of time (Spoorenberg 2016; also Johansson and Nygren 1991). Furthermore, while sex ratios tend to respond to secular changes in major demographic forces of fertility and mortality, the overwhelming majority of our regional populations (290/316) had not experienced the onset of a monotonic fertility decline and may well be taken to represent the demographic *ancien regime*[4].

---

[4] In order to assess whether particular populations captured in our database precede the onset of a monotonic fertility decline, our regional data were matched (by means of a spatial joint) to the corresponding province-level estimates of the onset of the fertility decline derived from the European Princeton Fertility Project’s capstone volume (Coale and Watkins 1986; for territories not covered by the Princeton data - such as Turkey, Albania, or Siberia - we used indicators derived from other existing literature, e.g. Falkingham and Gjonca 2001; or Coale et al. 1979, on Asiatic Russia).

B. MAIN CHARACTERISTICS OF THE REGIONAL POPULATIONS

This section reports the observed child sex ratios (CSR), as well as the results from bootstrapping these samples: upper and lower bound estimates (Figure 1B; for methodology, see the main text). Table 1B also provides the predicted value from our model (benchmark), as well as the model residuals, for all locations in our dataset. Lastly, the number of children in each region is provided, as well as how each region has been classified according census quality. The number of children is provided both as weighted (w.) and unweighted (unw.) figure. Census quality is classified as: 0-Modern Census, 1-Pre-modern censuses, and 2-Other censuses. Dark grey identifies locations whose sex ratio is significantly higher than the value predicted by the model (the lower bound interval estimate is higher than the predicted sex ratios).

Please note that due to sampling procedures, for some of the censuses the number of children in age group 0-4 years is available in both unweighted and weighted numbers. In the analysis the following principles were followed:

- the **unweighted numbers were used** to determine the N in the bootstrapping procedure;
- the **unweighted numbers were used** as the weights in GLM model.

In both cases the logic was that what matters is the original sample size - how many children "produced" the value of a given CSR, and/or how many "trials" resulted in the proportion of sexes used in the model

However, the CSRs are calculated from the **weighted numbers**, due to the fact that in some locations the samples were not random (more of quota samples) - this produced odd CSRs as different socio-economic or religious groups were not adequately represented in the original sample.

FIGURE 1B: Bootstrapped sex ratios and their corresponding 95% confidence intervals based on resampling with replacement (5,000 times) by sample size

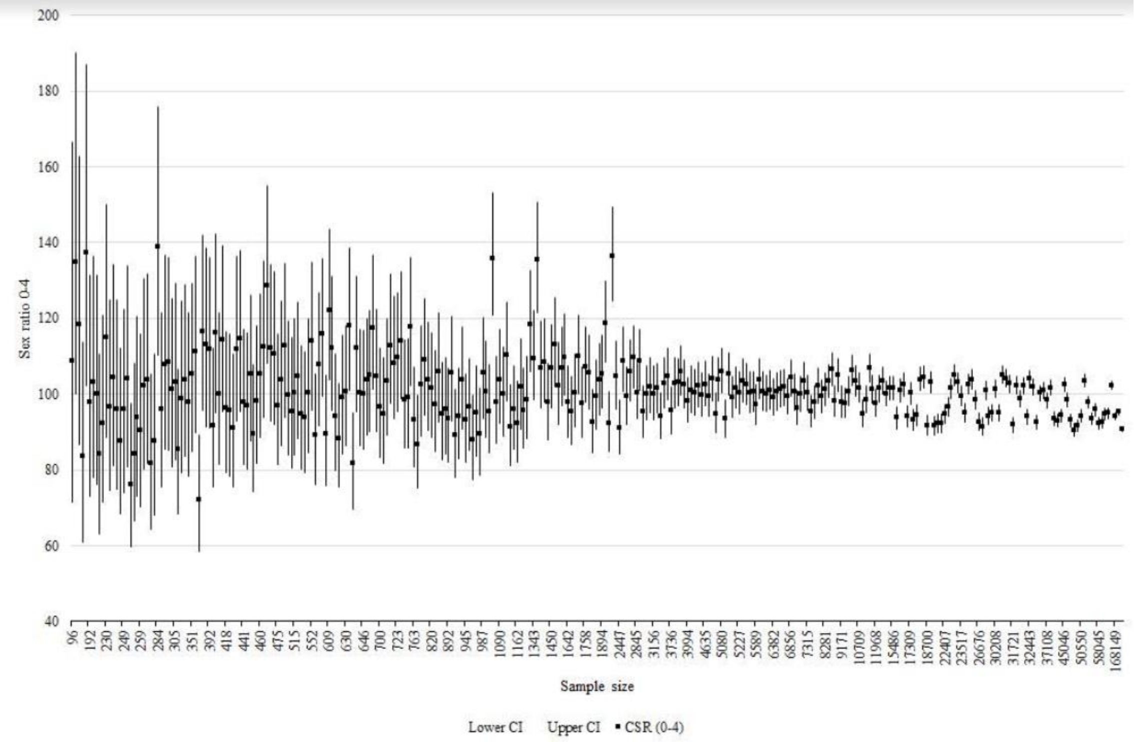


TABLE 1B

Region	Lower CI	CSR (0-4)	Upper CI	Pred CSR	Residual	Children 0-4 (unw.)	Children 0-4 (w.)	Census quality
AL/Berat 1918	121,0	135,6	153,1	103,5	32,1	1063	3501	0
AL/Durresi (city) 1918	95,7	116,5	142,1	96,0	20,5	368	368	0
AL/Elbasani (city) 1918	98,7	109,5	122,0	104,8	4,7	1343	1343	0
AL/Gora 1918	84,6	95,5	107,8	104,3	-8,8	1043	1922	0
AL/Kavaja (city) 1918	100,9	118,2	138,7	98,1	20,1	635	635	0
AL/Kruja (city) 1918	86,5	103,8	124,7	105,0	-1,2	483	483	0
AL/Kruja 1918	99,3	115,9	135,9	104,3	11,6	592	6760	0
AL/Puka 1918	101,5	117,4	136,6	104,2	13,1	691	4388	0
AL/Shkodra (city) 1918	101,1	108,8	117,1	99,9	8,9	2881	2881	0
AL/Shkodra 1918	100,0	109,9	120,9	102,9	7,0	1730	8672	0
AL/Tirana (city) 1918	106,0	118,5	132,6	100,5	18,0	1300	1300	0
AL/Tirana North 1918	124,6	136,4	149,4	102,3	34,1	2015	11236	0
AL/Tirana South 1918	121,6	135,3	150,6	104,7	30,6	1376	3550	0
AL/Zhuri 1918	108,4	118,6	129,8	103,8	14,8	1926	10715	0
AT/Styria 1910	90,4	103,8	119,1	109,5	-5,8	815	815	0
AT/Tyrol 1910	97,5	112,9	131,6	111,0	1,9	711	711	0
AT/Upper Austria 1910	100,0	134,9	190,2	103,6	31,4	148	148	0
AT/Waidhofen/Ybbs (city) 1910	81,6	100,0	121,4	108,1	-8,1	414	414	0
BE/Western Flanders 1814	90,7	99,5	109,0	92,8	6,6	1835	1835	1
BG/Čepelare (city) 1880-1947	73,0	97,9	131,3	105,1	-7,1	192	192	2
BG/Rhodope region 1877-1947	80,8	93,2	107,3	106,2	-13,0	763	763	2
CH/Zürich (city) 1870	76,3	89,4	105,2	105,6	-16,2	587	587	0
CH/Zürich North rural 1671-1685	85,7	98,7	114,2	104,4	-5,8	739	739	2
CH/Zürich South rural 1678-1762	79,1	93,9	111,3	105,2	-11,3	541	541	2
DE/Arnsberg 1846	80,9	104,0	133,9	97,3	6,7	255	224	1
DE/Braunschweig 1846	102,3	137,3	187,1	93,8	43,5	178	157	1
DE/Constance 1749-1811	64,2	81,9	105,3	100,0	-18,2	271	271	2
DE/Danzig and Posen 1858	88,5	105,4	126,6	89,3	16,1	460	460	1
DE/Düsseldorf 1846	76,1	100,0	131,5	90,2	9,8	206	181	1
DE/Höhscheid 1846	82,8	94,2	107,1	96,2	-2,0	938	938	1
DE/Koblenz 1846	75,6	95,9	121,5	96,8	-0,9	288	253	1
DE/Liegnitz 1846	66,5	84,3	108,1	91,9	-7,6	258	227	1
DE/Lippborg 1861	81,1	104,3	134,3	95,5	8,7	239	239	1
DE/Mecklenburg-Schwerin Northeast 1819	87,1	97,8	110,2	99,3	-1,5	1076	1076	0
DE/Mecklenburg-Schwerin Northeast 1867	96,4	107,0	119,2	98,8	8,2	1416	1416	0
DE/Mecklenburg-Schwerin Northwest 1819	80,3	102,3	130,4	101,6	0,7	265	265	0
DE/Mecklenburg-Schwerin Northwest 1867	81,3	97,1	115,9	102,2	-5,1	475	475	0
DE/Mecklenburg-Schwerin Southeast 1867	78,3	95,8	116,0	101,3	-5,5	419	419	0
DE/Mecklenburg-Schwerin Southwest 1819	88,1	97,8	108,0	101,1	-3,2	1450	1450	0
DE/Mecklenburg-Schwerin Southwest 1867	88,6	97,5	107,3	101,4	-3,9	1758	1758	0
DE/Meppen 1749	88,3	98,4	110,0	96,4	2,0	1256	1256	2
DE/Meppen and Cloppenburg 1700	92,9	105,8	120,4	96,9	8,9	928	928	2
DE/Merseburg 1846	70,4	90,4	115,8	90,7	-0,2	259	228	1

DE/Münster 1846	71,4	108,7	166,7	90,1	18,6	96	84	1
DE/Rheine-Bevergern 1700	79,3	98,7	124,5	95,7	3,1	312	312	2
DE/Rheine-Bevergern 1749	69,8	81,8	95,1	96,7	-14,9	640	640	2
DE/Rostock (city) 1867	101,7	109,6	117,9	99,9	9,7	2783	2783	0
DE/Sachsen-Coburg 1846	68,1	87,7	110,8	93,6	-5,9	274	241	1
DE/Sachsen-Gotha 1846	74,8	96,0	124,9	94,2	1,7	243	214	1
DE/Sigmaringen 1861	81,7	94,8	109,8	91,9	2,9	705	705	1
DE/Stromberg 1749	96,3	105,4	115,5	99,5	6,0	1894	1894	2
DE/Trier 1846	88,5	115,0	150,0	96,3	18,7	230	202	1
DE/urban Centre 1846	68,5	85,6	106,7	92,2	-6,6	310	214	1
DE/urban East 1846	95,1	116,1	142,2	87,5	28,6	402	277	1
DE/urban West 1846	95,3	114,4	139,1	92,7	21,7	416	287	1
DE/Vechta 1700	83,2	96,6	112,1	97,1	-0,5	700	700	2
Denmark/Aalborg 1787	95,1	100,9	107,5	102,1	-1,3	4291	4291	2
Denmark/Aarhus 1787	88,7	95,3	102,3	104,1	-8,8	2889	2889	2
Denmark/Frederiksborg 1787	99,6	105,3	111,0	103,0	2,3	5095	5095	2
Denmark/Hjorring 1787	96,9	102,7	108,7	102,6	0,1	4635	4635	2
Denmark/Kobenhavn 1787	94,3	97,7	101,2	100,4	-2,7	11968	11968	2
Denmark/Maribo 1787	95,6	100,7	105,9	98,4	2,3	6010	6010	2
Denmark/Odense 1787	98,8	103,5	108,5	102,4	1,1	7281	7281	2
Denmark/Prasto 1787	96,1	101,0	106,2	102,4	-1,4	6335	6335	2
Denmark/Randers 1787	88,5	93,5	98,5	102,8	-9,4	5088	5088	2
Denmark/Ribe 1787	92,6	98,3	104,5	101,8	-3,5	3994	3994	2
Denmark/Ringkobing 1787	96,4	102,6	109,2	101,2	1,4	3989	3989	2
Denmark/Roskilde 1787	92,0	99,3	107,2	102,9	-3,6	2629	2629	2
Denmark/Soro 1787	89,7	94,9	100,5	101,3	-6,4	4922	4922	2
Denmark/Thisted 1787	93,1	100,0	107,4	101,8	-1,8	2966	2966	2
Denmark/Vejle 1787	94,2	99,6	105,5	103,8	-4,2	4668	4668	2
Denmark/Viborg 1787	96,7	102,3	108,3	102,6	-0,3	4581	4581	2
England/Bedfordshire 1851	91,6	94,5	97,3	93,5	1,0	17431	17431	1
England/Berkshire 1851	89,2	91,8	94,5	92,8	-1,1	18700	18700	1
England/Buckinghamshire 1851	90,9	93,7	96,6	93,8	-0,1	16087	16087	1
England/Cambridgeshire 1851	92,6	95,0	97,5	91,2	3,8	23722	23722	1
England/Cheshire 1851	94,3	95,9	97,6	94,9	1,0	52775	52775	1
England/Cornwall 1851	96,7	98,5	100,3	96,9	1,5	45567	45567	1
England/Cumberland 1851	96,1	98,6	101,1	99,2	-0,6	24384	24384	1
England/Derbyshire 1851	90,0	92,0	94,0	97,5	-5,5	31721	31721	1
England/Devonshire 1851	93,3	94,8	96,2	97,7	-2,9	67248	67248	1
England/Dorset 1851	89,1	91,6	94,3	95,8	-4,2	21173	21173	1
England/Durham 1851	96,2	97,8	99,5	96,0	1,8	51621	51621	1
England/Essex 1851	91,3	93,0	94,8	93,4	-0,4	41371	41371	1
England/Gloucestershire 1851	91,5	93,1	94,8	95,5	-2,4	46385	46385	1
England/Hampshire 1851	88,9	90,5	92,2	95,0	-4,4	47307	47307	1
England/Herefordshire 1851	91,5	94,8	98,4	96,7	-1,9	11266	11266	1
England/Hertfordshire 1851	89,9	92,3	94,7	93,7	-1,4	22273	22273	1
England/Huntingdonshire 1851	91,2	95,4	99,9	90,8	4,6	7615	7615	1

England/Kent (extra London) 1851	91,1	92,6	94,1	95,7	-3,1	59537	59537	1
England/Lancashire 1851	94,5	95,3	96,1	96,1	-0,8	250781	250781	1
England/Leicestershire 1851	92,9	95,0	97,3	94,6	0,4	29470	29470	1
England/Lincolnshire 1851	92,7	94,3	95,9	92,5	1,8	50550	50550	1
England/London (Parts Of Middlesex, Surrey and Kent) 1851	90,0	90,7	91,4	94,1	-3,5	268797	268797	1
England/Middlesex (extra London) 1851	91,3	94,1	97,0	93,4	0,7	17219	17219	1
England/Norfolk 1851	91,9	93,6	95,2	91,8	1,7	51637	51637	1
England/Northamptonshire 1851	89,2	91,4	93,7	93,1	-1,7	26887	26887	1
England/Northumberland 1851	96,4	98,4	100,4	96,9	1,5	37108	37108	1
England/Nottinghamshire 1851	90,6	92,6	94,6	93,0	-0,4	33458	33458	1
England/Oxford 1851	90,6	93,3	96,1	92,8	0,5	17355	17355	1
England/Rutland 1851	93,5	100,5	108,1	93,5	7,0	2845	2845	1
England/Shropshire 1851	92,1	94,3	96,5	96,7	-2,5	28501	28501	1
England/Somerset 1851	90,0	91,6	93,3	96,7	-5,1	50082	50082	1
England/Staffordshire 1851	93,7	95,0	96,2	95,2	-0,2	85922	85922	1
England/Suffolk 1851	92,6	94,4	96,2	92,4	2,1	42605	42605	1
England/Surrey (extra London) 1851	89,8	92,2	94,7	95,2	-3,0	21641	21641	1
England/Sussex 1851	91,7	93,5	95,3	96,4	-2,8	41048	41048	1
England/Warwick 1851	90,9	92,4	93,9	93,3	-0,9	58045	58045	1
England/Westmorland 1851	95,9	100,3	105,2	100,4	-0,1	7315	7315	1
England/Wiltshire 1851	90,5	92,8	95,0	94,7	-2,0	26676	26676	1
England/Worcestershire 1851	92,0	94,1	96,2	94,4	-0,3	32383	32383	1
England/Yorkshire East Riding 1851	92,9	95,0	97,2	96,2	-1,1	30350	30350	1
England/Yorkshire North Riding 1851	92,2	94,6	97,2	97,2	-2,5	22407	22407	1
England/Yorkshire West Riding 1851	93,2	94,1	95,0	96,6	-2,6	168149	168149	1
ES/Andalucía 1752	68,3	87,7	112,2	96,5	-8,8	244	244	1
ES/Barcelona province urban 1889	98,9	114,1	132,5	103,5	10,5	730	730	0
ES/Catalonia rural 1880-1890	102,1	117,8	136,0	105,8	12,0	760	760	0
ES/other Catalonia urban 1889	92,2	103,7	117,1	103,8	0,0	1090	1090	0
F/Northeast 1846	81,7	98,3	118,2	95,2	3,0	456	456	1
F/Northwest 1846	84,3	99,0	115,6	94,0	5,0	623	623	1
F/South 1846	103,7	122,3	143,6	97,7	24,5	609	609	1
F/Southwest 1831-1901	75,5	91,1	110,8	97,6	-6,5	430	430	1
F/St. Emilion (city) 1846/1856	75,7	91,8	112,3	96,8	-5,0	397	397	1
GR/Kythira 1724	86,4	100,6	118,0	104,9	-4,3	630	630	2
H/Great Plain 1869	83,8	99,6	119,3	90,3	9,4	489	468	0
H/North-East 1869	110,4	138,8	175,7	97,9	40,9	284	268	0
H/Northern Transdanubia 1869	80,4	95,5	114,8	98,2	-2,6	507	511	0
H/Southern Transdanubia 1869	79,4	96,2	116,6	94,5	1,7	418	417	0
HR/Dubrovnik area 1674	74,6	96,6	125,0	105,1	-8,5	234	234	2
Iceland 1703	89,6	95,6	101,8	94,0	1,6	3736	3736	1
LT/Kovno 1847	90,9	103,7	117,8	96,0	7,7	941	941	2
LT/Vilna 1847	85,5	95,9	107,6	99,2	-3,3	1150	1150	2
LV/Courland Goldingen+Pilten 1797	84,6	92,6	101,2	100,8	-8,2	1831	18820	2
LV/Courland Mitau 1797	92,1	102,3	113,6	99,9	2,3	1506	15778	2
LV/Courland Selburg 1797	83,6	95,0	107,6	99,8	-4,8	984	11498	2

LV/Courland Tuckum 1797	93,5	108,2	125,9	100,0	8,2	714	8333	2
NL/Eindhoven and Helmond (city) 1811	80,3	94,9	112,9	87,4	7,5	530	530	1
NL/Goes (city) 1810	83,9	100,4	120,1	83,7	16,7	515	515	1
NL/North Brabant 1810	85,6	100,0	116,8	88,5	11,5	646	646	1
NL/Tilburg (city) 1811	90,8	102,1	114,7	89,1	13,0	1166	1166	1
NL/Zeeland 1811	91,4	100,5	110,7	84,5	16,0	1700	1700	1
Norway/Aggershuus 1801	93,5	97,8	102,4	96,8	1,1	7648	7648	1
Norway/Bratsberg 1801	94,3	99,2	104,2	99,4	-0,2	6382	6382	1
Norway/Buskerud 1801	97,6	102,2	106,8	99,5	2,7	7856	7856	1
Norway/Christiania 1801	82,0	93,5	107,0	95,9	-2,4	892	892	1
Norway/Christians 1801	93,9	97,6	101,7	101,2	-3,5	9588	9588	1
Norway/Finmarken 1801	96,4	102,9	109,8	100,6	2,3	3756	3756	1
Norway/Hedemarken 1801	95,2	99,4	103,9	98,8	0,6	7970	7970	1
Norway/Jarlsberg O 1801	94,9	100,3	105,9	98,6	1,8	5227	5227	1
Norway/Lister Og M 1801	98,2	103,7	109,8	99,2	4,5	5060	5060	1
Norway/Nedenas 1801	96,9	103,0	109,5	98,6	4,4	3939	3939	1
Norway/Nordland 1801	96,5	101,2	106,2	102,2	-1,0	6718	6718	1
Norway/Nordre Berg 1801	95,4	100,2	105,1	102,7	-2,5	6246	6246	1
Norway/Nordre Tron 1801	95,2	100,4	105,9	100,2	0,2	5523	5523	1
Norway/Romsdal 1801	94,8	99,4	104,2	100,8	-1,5	6832	6832	1
Norway/Smaalehnene 1801	91,9	96,3	100,9	95,3	0,9	7018	7018	1
Norway/Sondre Berg 1801	96,8	100,7	104,8	101,3	-0,6	9724	9724	1
Norway/Sondre Tron 1801	99,5	104,3	109,2	99,0	5,3	6856	6856	1
Norway/Stavanger 1801	93,7	99,0	104,7	99,5	-0,4	5111	5111	1
Norway/Trondheim 1801	88,7	100,8	114,1	98,9	2,0	1002	1002	1
PL/Central Belarus 1768-1804	84,3	91,2	98,6	94,8	-3,6	2447	2447	2
PL/Chelmska Land 1791-1792	93,4	100,1	107,5	94,8	5,3	3156	3156	2
PL/Greater Poland 1666-1809	95,5	112,2	131,2	95,8	16,4	645	645	2
PL/Kujavia 1766-1792	97,8	107,1	117,8	96,0	11,1	1758	1758	2
PL/Lesser Poland 1789-1792	84,7	92,3	100,6	97,6	-5,3	2008	2008	2
PL/Ostrzeszow County 1790-1791	82,1	92,4	103,5	96,1	-3,7	1162	1162	2
PL/Podolia 1785-1819	86,4	100,3	117,2	103,4	-3,1	645	645	2
PL/Polesia 1795	88,2	94,1	100,4	93,6	0,5	3673	3673	2
PL/Silesian lowlands 1792	58,5	72,0	89,1	98,2	-26,1	363	363	2
PL/Sudenten 1805	82,6	103,3	129,3	104,0	-0,7	305	305	2
PL/Sudeten 1781-82	79,9	94,0	110,5	102,2	-8,1	619	619	2
PL/Warmia 1695-1772	81,0	101,3	125,2	97,3	4,1	304	304	2
PL/Wielunskie County 1790-1792	81,2	91,2	102,5	97,0	-5,8	1136	1136	2
PL/Zhytomyr County 1791	96,0	105,6	115,6	96,5	9,0	1766	1766	2
RO/Eastern Wallachia 1838	85,1	96,5	109,4	94,3	2,2	957	957	1
RO/Moldavia Catholics 1781-1787	91,4	113,0	138,6	98,5	14,5	377	377	2
RO/Moldavia Catholics 1866-1879	94,0	112,8	134,5	96,8	16,0	483	483	2
RO/Northern Wallachia 1838	77,9	89,1	101,3	96,0	-6,9	934	934	1
RO/Partium 1869	91,2	111,9	136,1	99,9	12,0	392	364	0
RO/Southern Wallachia 1838	77,3	88,1	99,8	92,6	-4,5	963	963	1
RO/Southwestern Wallachia 1838	78,6	89,6	101,6	93,4	-3,8	984	984	1

RO/Transylvania 1869	90,1	104,7	122,5	100,3	4,4	692	730	0
RUS/Braclav Governorate 1795	98,2	110,4	124,2	102,0	8,4	1130	1130	2
RUS/Cherdyn (city) 1710	73,2	94,0	120,5	96,0	-2,0	258	258	2
RUS/Chusovoy 1710	98,1	106,0	114,3	94,7	11,3	2651	2651	2
RUS/Gagarin villages 1814	95,6	114,6	137,8	95,2	19,4	440	440	2
RUS/Ilyinsky 1710	96,5	106,8	118,4	94,6	12,3	1450	1450	2
RUS/Ilyinsky West 1710	88,8	102,6	118,0	93,8	8,8	774	774	2
RUS/Kama West 1710	89,3	103,4	119,9	93,4	10,1	708	708	2
RUS/Kaygorodok 1710	92,8	111,8	136,3	96,7	15,2	430	430	2
RUS/Moscow area 1897	88,7	97,8	107,6	96,1	1,7	1642	1642	0
RUS/Solikamsk (city) 1710	85,8	99,2	114,7	96,5	2,7	745	745	2
RUS/Solikamsk Center 1710	82,1	93,3	105,4	95,7	-2,5	945	945	2
RUS/Solikamsk North 1710	86,8	95,4	104,6	94,5	0,8	1696	1696	2
RUS/Solikamsk South 1710	97,8	108,4	119,9	95,4	13,0	1434	1434	2
RUS/Southern Perm 1710	93,7	112,4	135,0	94,3	18,1	463	463	2
RUS/Southwest Perm 1710	96,8	107,1	117,8	94,5	12,6	1555	1555	2
RUS/St. Petersburg area South 1811	84,7	100,4	119,8	93,2	7,2	543	543	2
RUS/St. Petersburg area West 1811	89,1	103,7	119,9	93,6	10,1	662	662	2
RUS/Tula (city) 1710	85,0	97,4	111,5	98,1	-0,8	825	825	2
RUS/Ural Centre 1710	84,3	95,9	109,7	94,3	1,6	868	868	2
RUS/Ural East 1710	80,1	96,9	116,1	93,8	3,1	443	443	2
RUS/Ural iron plants 1710	81,8	103,8	131,9	97,1	6,7	269	269	2
RUS/Ural North 1710	75,3	86,7	99,7	93,9	-7,2	773	773	2
RUS/Ural North Centre 1710	108,1	128,6	154,9	93,9	34,6	464	464	2
RUS/Ural South 1710	88,5	101,5	116,4	93,6	7,9	820	820	2
RUS/Ural Verkhoturye (city) 1710	74,1	96,1	122,3	94,6	1,5	249	249	2
RUS/Ural West 1710	90,2	105,2	122,0	94,7	10,4	677	677	2
RUS/Vyatka (city) 1710	91,6	107,7	126,9	99,4	8,3	590	590	2
RUS/West Siberia 1710	93,5	105,6	120,3	99,8	5,8	987	987	2
RUS/Yarensk East 1710	59,6	76,0	97,7	97,7	-21,7	257	257	2
RUS/Yarensk West 1710	74,2	89,5	107,8	95,3	-5,8	453	453	2
Scotland/Aberdeenshire 1881	99,7	101,7	103,8	104,2	-2,5	37903	37903	0
Scotland/Angus and Forfarshire 1881	98,8	100,9	103,0	101,3	-0,4	35624	35624	0
Scotland/Argyll 1881	100,8	105,1	109,5	106,0	-0,9	9080	9080	0
Scotland/Ayr 1881	100,0	102,2	104,5	102,0	0,2	32176	32176	0
Scotland/Banff 1881	99,1	103,4	107,9	107,1	-3,7	8675	8675	0
Scotland/Berwickshire 1881	94,6	100,4	106,5	103,7	-3,2	4498	4498	0
Scotland/Bute 1881	96,1	104,7	114,0	107,0	-2,3	2037	2037	0
Scotland/Caithness 1881	94,2	99,8	105,8	103,9	-4,1	4601	4601	0
Scotland/Clackmannanshire 1881	95,0	101,5	108,2	106,7	-5,2	3615	3615	0
Scotland/Dumfriesshire 1881	97,0	101,2	105,6	104,3	-3,1	8281	8281	0
Scotland/Dunbartonshire 1881	99,5	103,4	107,4	102,8	0,6	10340	10340	0
Scotland/East Lothian and Haddingtonshire 1881	100,4	106,1	112,1	104,2	1,9	5080	5080	0
Scotland/Fife 1881	102,3	105,1	107,8	103,3	1,8	23006	23006	0
Scotland/Inverness-shire 1881	102,2	106,2	110,4	106,1	0,2	10233	10233	0
Scotland/Kincardinshire 1881	98,3	104,0	110,0	104,5	-0,5	4788	4788	0

Scotland/Kinross-shire 1881	82,6	94,7	108,6	104,2	-9,5	851	851	0
Scotland/Kirkudbrightshire 1881	95,3	100,7	106,1	104,5	-3,8	5551	5551	0
Scotland/Lanarkshire 1881	101,3	102,4	103,5	102,6	-0,2	137035	137035	0
Scotland/Mid Lothian and Edinburghshire 1881	101,8	103,5	105,4	103,3	0,2	51246	51246	0
Scotland/Morayshire and Elginshire 1881	98,8	103,9	109,3	102,9	1,0	5848	5848	0
Scotland/Nairnshire 1881	88,9	100,0	112,5	104,0	-4,0	1118	1118	0
Scotland/Orkney 1881	98,6	104,8	112,1	105,2	-0,4	3684	3684	0
Scotland/Peeblesshire 1881	94,6	103,7	113,6	106,8	-3,1	1841	1841	0
Scotland/Perthshire 1881	97,0	100,2	103,5	106,2	-6,0	15205	15205	0
Scotland/Renfrewshire 1881	101,9	104,2	106,6	102,7	1,5	31151	31151	0
Scotland/Ross and Cromarty 1881	102,1	106,5	111,0	106,2	0,3	8919	8919	0
Scotland/Roxburghshire 1881	95,8	100,6	105,4	105,3	-4,7	7011	7011	0
Scotland/Selkirk 1881	96,4	102,8	109,8	104,9	-2,1	3676	3676	0
Scotland/Shetland 1881	95,0	102,0	109,6	103,5	-1,5	2999	2999	0
Scotland/Stirlingshire 1881	99,4	102,5	105,7	105,0	-2,5	16966	16966	0
Scotland/Sutherland 1881	100,7	108,7	117,6	104,5	4,2	2481	2481	0
Scotland/West Lothian and Linlithgowshire 1881	97,0	101,8	106,7	104,2	-2,4	6825	6825	0
Scotland/Wigtown 1881	96,3	101,7	107,3	102,8	-1,1	5201	5201	0
SE/Belgrade 1733	60,8	83,5	113,7	102,2	-18,6	156	156	2
SE/Jasenica 1863	85,7	95,6	107,0	95,4	0,2	1209	1209	1
SE/Jasenica 1884	99,6	109,6	121,3	101,1	8,5	1587	1587	0
SE/Kruševac (city) 1863	84,7	105,3	129,4	93,2	12,0	351	351	1
SK/Central 1869	63,2	84,3	110,6	103,7	-19,4	217	234	0
SK/East 1869	91,9	110,5	132,4	103,0	7,5	474	471	0
SK/West 1869	89,9	111,3	136,4	98,2	13,2	357	374	0
Sweden/Älvsborg 1880	101,7	104,0	106,3	98,3	5,7	32443	32443	1
Sweden/Blekinge 1880	101,5	104,5	107,5	96,6	8,0	18582	18582	1
Sweden/Gävleborg 1880	101,3	103,9	106,5	98,8	5,0	24177	24177	1
Sweden/Göteborg och Bohus 1880	100,5	102,7	105,0	98,8	3,9	31438	31438	1
Sweden/Gotland 1880	97,9	103,5	109,3	95,7	7,8	5315	5315	1
Sweden/Halland 1880	97,4	100,3	103,3	98,1	2,2	17309	17309	1
Sweden/Jämtland 1880	103,0	106,8	110,8	100,5	6,3	11678	11678	1
Sweden/Jönköping 1880	99,0	101,6	104,2	98,0	3,6	22751	22751	1
Sweden/Kalmar 1880	102,7	105,1	107,5	97,5	7,6	30636	30636	1
Sweden/Kopparberg 1880	100,0	102,6	105,3	98,5	4,1	24154	24154	1
Sweden/Kristianstad 1880	98,5	101,0	103,4	97,7	3,2	26930	26930	1
Sweden/Kronoberg 1880	100,3	103,1	106,0	96,3	6,8	20326	20326	1
Sweden/Malmöhus 1880	100,7	102,5	104,4	96,7	5,9	45046	45046	1
Sweden/Norrbotten 1880	100,0	103,5	107,2	98,1	5,5	13419	13419	1
Sweden/Örebro 1880	100,5	103,1	105,7	97,6	5,5	23172	23172	1
Sweden/Östergötland 1880	99,9	102,0	104,2	98,2	3,8	32539	32539	1
Sweden/Skaraborg 1880	99,1	101,4	103,7	97,3	4,0	30208	30208	1
Sweden/Södermanland 1880	100,9	103,9	107,1	98,5	5,4	17440	17440	1
Sweden/Stockholm 1880	98,1	100,3	102,4	96,5	3,7	33629	33629	1
Sweden/Uppsala 1880	97,8	101,1	104,2	96,9	4,1	16262	16262	1
Sweden/Värmland 1880	100,2	102,4	104,7	100,2	2,1	32026	32026	1

Sweden/Västerbotten 1880	98,4	101,6	104,8	97,1	4,5	15486	15486	1
Sweden/Västernorrland 1880	97,0	99,5	102,1	99,5	0,0	23517	23517	1
Sweden/Västmanland 1880	98,4	101,6	104,8	96,4	5,2	15388	15388	1
TR/Istanbul (city) 1885	83,7	103,9	129,0	102,7	1,2	316	316	0
TR/Istanbul (city) 1907	81,5	97,8	117,2	103,8	-6,0	441	441	0
UKR/Berdychiv (city) 1897	102,1	113,2	125,5	98,8	14,3	1475	4425	0
UKR/Berdychiv region North rural 1897	88,3	104,7	124,3	98,5	6,2	516	12402	0
UKR/Berdychiv region North urban 1897	78,3	97,9	121,6	100,6	-2,7	328	2874	0
UKR/Berdychiv region South rural 1897	85,3	108,6	135,9	100,8	7,8	302	7710	0
UKR/Berdychiv region South urban 1897	87,5	105,3	126,1	101,0	4,4	450	3456	0
UKR/Hetmanate North 1765	75,4	88,2	102,9	97,4	-9,2	621	621	2
UKR/Hetmanate Southeast 1765	92,7	105,9	121,5	98,3	7,6	844	844	2
UKR/Hetmanate Southwest 1765	94,5	109,0	125,2	96,3	12,7	813	813	2
UKR/Hetmanate West 1765	75,9	89,7	105,1	95,7	-6,0	605	605	2
USSR/Central 1926/27	95,9	112,1	131,2	99,2	12,8	615	615	0
USSR/East 1926/27	71,5	92,2	120,8	102,4	-10,2	223	223	0
USSR/North 1926/27	95,7	114,0	134,9	98,8	15,1	552	552	0
USSR/Northeast 1926/27	85,4	107,9	136,6	100,4	7,5	291	291	0
USSR/Northwest 1926/27	77,9	103,0	136,5	101,6	1,5	201	201	0
USSR/South 1926/27	93,0	112,2	134,3	101,6	10,5	471	471	0
USSR/Southeast 1926/27	86,7	118,3	162,7	103,4	14,9	155	155	0
USSR/West 1926/27	95,4	109,6	126,6	98,7	10,8	723	723	0
Wales/Anglesey 1851	92,1	97,2	102,5	96,2	1,0	5589	5589	1
Wales/Brecknockshire 1851	95,5	100,0	104,7	98,2	1,8	7213	7213	1
Wales/Cardiganshire 1851	98,0	101,5	105,1	101,3	0,2	12472	12472	1
Wales/Carmarthenshire 1851	95,1	98,5	102,1	98,5	0,0	11598	11598	1
Wales/Carnarvonshire 1851	97,7	101,3	105,0	98,0	3,3	11748	11748	1
Wales/Denbighshire 1851	98,0	101,8	105,7	98,3	3,4	10709	10709	1
Wales/Flintshire 1851	97,2	102,5	108,1	97,6	4,9	5323	5323	1
Wales/Glamorganshire 1851	96,7	98,8	101,0	98,7	0,1	32068	32068	1
Wales/Merionethshire 1851	95,7	100,7	105,6	101,3	-0,6	6385	6385	1
Wales/Monmouthshire 1851	94,1	96,6	99,2	98,4	-1,8	22679	22679	1
Wales/Montgomeryshire 1851	94,1	98,1	102,1	99,8	-1,6	9034	9034	1
Wales/Pembrokeshire 1851	94,0	98,0	102,2	97,0	1,1	9171	9171	1
Wales/Radnorshire 1851	99,7	106,1	112,9	98,7	7,4	3953	3953	1

Note: census quality is denoted by – “0” (Modern state census), “1” (Premodern state census), and “2” (Other census).

C. DATASET ON HISTORICAL INFANT MORTALITY RATES

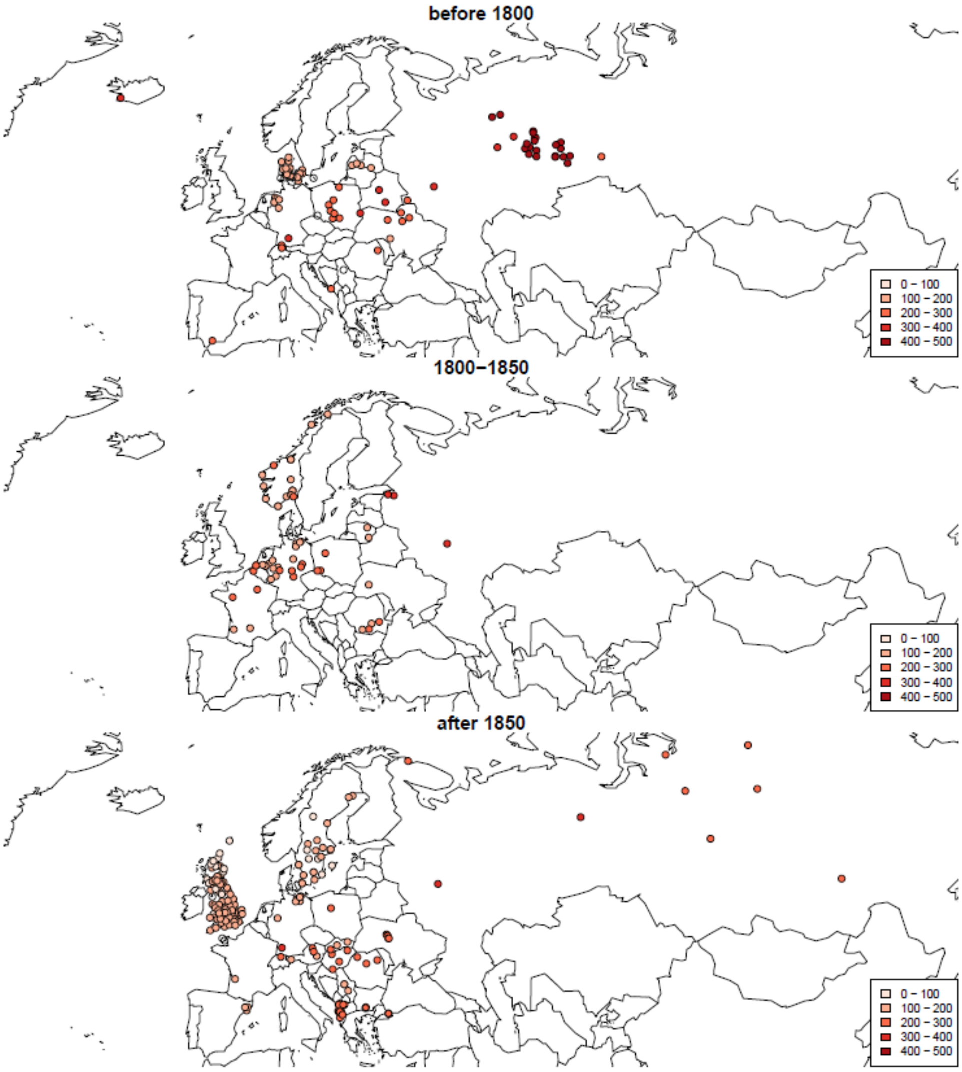
We have collected a novel set of information on the prevailing infant mortality rates (henceforth, IMR) for nearly all our locations (308/316). Given the lack of *omnibus*, harmonized, temporally sensitive, large-scale and high-resolution evidence on the variability of IMR in the European past (cf. Kluesener et al. for 1910), our figures are based on available regional statistics and the voluminous secondary literature.

High-quality IMR data corresponding both geographically and temporally to the spatial units contained in the NAPP/Mosaic data were extremely rare. Such data were nevertheless avaliable for England, Wales and Scotland, Sweden, most of the Flemish and Dutch data, as well as parts of Denmark. For the remaining locations, the data collection involved many challenges which need to be born in mind when interpreting this information:

- *Data fusion and linkage across scales*: for some regions, the available IMR data came from small number of individual parishes contained within those regions; for others, information could only be obtained from larger administrative units to which these regions belonged; yet in some other cases, the respective IMRs could only be established by extrapolation from detailed case studies of neighbouring populations (usually close chronological resemblance supported such a choice);
- *Time discrepancy*: except for GB, Sweden, Norway, NL and Belgium, as well as some parts of Germany, for the majority of our locations the IMR data could be found for dates only roughly approximating a particular census date, and most often for time periods subsequent to the census;
- *Territorial overlay of multiple spatial units*: in some instances, our regions consisted of populations spread across administrative units with different values of infant mortality. This required averaging the available IMR data over various units; however, having no access to underlying detailed population statistics, this usually proceeded without the application of any weights (see Table B1 for details)

Figure 1C depicts the information that has been gathered geographically and with a broad temporal breakdown, and Table 1C provides the basic descriptive statistics for these three time periods. Table 2C presents all collected data in a regional breakdown, including references to the sources of information. We are of the opinion that despite the drawbacks described above, the overall picture of the distribution of the IMRs across our dataset appears to be pretty consistent with what is to be expected from the literature on the evolution of IMR in historical Europe (e.g. Corsini and Viazzo 1997).

FIGURE 1C: Distribution of the NAPP/Mosaic IMR data by three main time periods



Source: NAPP/Mosaic Historical Infant Moratlity Datset.

TABLE 1C: Descriptive statistics for the NAPP/Mosaic IMR data by time periods

before 1800	N	Valid	75
		Missing	7
	Mean		276
	Median		241
1800-1850	N	Valid	57
		Missing	0
	Mean		195
	Median		185
after 1850	N	Valid	172
		Missing	5
	Mean		154
	Median		141

Source: NAPP/Mosaic Historical Infant Moratlity Datset.

TABLE 2C: Descriptive information on the collected NAPP/Mosaic IMR data

Region1	***IMR***	***Area referred to***	***Time period***	***Source***
AL/Berat 1918	**200**	the parish of Shkodra	Around 1900	Catholic parish registers, Gruber, personal infor.
AL/Durresi (city) 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Elbasani (city) 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Gora 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Kavaja (city) 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Kruja (city) 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Kruja 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Puka 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Shkodra (city) 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Shkodra 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Tirana (city) 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Tirana North 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Tirana South 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AL/Zhuri 1918	**200**	the parish of Shkodra	Around 1900	Ibid.
AT/Styria 1910	**199**	exact	1906-10	Kytir, Köck and Münz 1995
AT/Tyrol 1910	**167**	exact	1906-10	Ibid.
AT/Upper Austria 1910	**221**	exact	1906-10	Ibid.
AT/Waidhofen/Ybbs (city) 1910	**221**	Province Lower Austria	1906-10	Ibid.
BE/Western Flanders 1814	**160**	pop. weighted average from IMRs for 4 out of 5 villages in the region	1800-1809	Ghent University, Quetelet Centre, STREAM dataset (infant mortality) and LOKSTAT dataset (population). Courtesy of I. Devos.

BG/Čepelare (city) 1880-1947	**251**	Seldzhikovo parish	1853-72	Todorova 2006, 85.
BG/Rhodope region 1877-1947	**161**	whole country	1910 https://www.statista.com	
CH/Zürich (city) 1870	**216**	exact city	1876-80	Fanconi 1933, p. 15,
CH/Zürich North rural 1671-1685	**202**	Maschwanden, ca. 50 kms from region's centroid	1709-49	Letsch 2017, p. 171
CH/Zürich South rural 1678-1762	**202**	Maschwanden, ca. 50 kms from region's centroid	1709-49	Ibid.
DE/Arnsberg 1846	**144**	Arnsberg Rgbzk (where it currently fits)	1862-66	Knodel 1974, 289
DE/Braunschweig 1846	**190**	exact	1875-77	Ibid.
DE/Constance 1749-1811	**322**	whole province of Konstanz	1864-70	Ibid.
DE/Danzig and Posen 1858	**238**	Aver from D and P provinces	1862-66	Ibid.
DE/Düsseldorf 1846	**146**	exact	1862-66	Ibid.
DE/Höhscheid 1846	**146**	Province Düsseldorf	1862-66	Ibid., 288
DE/Koblenz 1846	**173**	exact province	1862-66	Ibid.
DE/Liegnitz 1846	**278**	exact	1862-66	Ibid., 289
DE/Lippborg 1861	**144**	Arnsberg Rgbzk (where it currently fits)	1862-66	Ibid., 289
DE/Mecklenburg-Schwerin Northeast 1819	**154**	Mecklenburg	1867-70	Ibid., 289
DE/Mecklenburg-Schwerin Northeast 1867	**158**	Mecklenburg-Schwerin province	1875-80	Ibid., 289
DE/Mecklenburg-Schwerin Northwest 1819	**154**	Mecklenburg	1867-70	Ibid., 289
DE/Mecklenburg-Schwerin Northwest 1867	**154**	Mecklenburg	1867-70	Ibid., 289

DE/Mecklenburg-Schwerin Southeast 1867	**154**	Mecklenburg	1867-70	Ibid., 289
DE/Mecklenburg-Schwerin Southwest 1819	**154**	Mecklenburg	1867-70	Ibid., 289
DE/Mecklenburg-Schwerin Southwest 1867	**154**	Mecklenburg	1867-70	Ibid., 289
DE/Meppen 1749	**186**	The parish of Hartum, some 140 kms from the border of the area	1700-49	Kloke 1998, p. 87
DE/Meppen and Cloppenburg 1700	**130**	The parish of Hesel, in the neighbourhood	1780 (?)	Mühlichen 2015
DE/Merseburg 1846	**219**	exact province	1862-66	Knodel 1974, 288
DE/Münster 1846	**153**	exact province	1862-66	Ibid., 289
DE/Rheine-Bevergern 1700	**186**	The parish of Hartum, some 130 kms from the border of the area	1700-49	Kloke 1998, p. 87
DE/Rheine-Bevergern 1749	**186**	The parish of Hartum, some 130 kms from the border of the area	1700-49	Kloke 1998, p. 87
DE/Rostock (city) 1867	**210**	exact (1 parish)	1865	Mühlichen 2015
DE/Sachsen-Coburg 1846	**218**	Thuringen province	1868-77	Knodel 1974, 288
DE/Sachsen-Gotha 1846	**218**	Thuringen province	1868-77	Knodel 1974, 289
DE/Sigmaringen 1861	**317**	exact province	1862-66	Ibid., 288
DE/Stromberg 1749	**186**	The parish of Hartum, some 75 kms from the border of the area	1700-49	Kloke 1998, p. 87
DE/Trier 1846	**154**	exact province	1862-66	Knodel 1974, 164, 289
DE/urban Centre 1846	**205**	Unwght average from IMRs from 3 provinces covered (Thuringen, Potsdam, Anhalt)	1862-77	Ibid.

DE/urban East 1846	**284**	Unwght average from 5 urban parishes in Greater Poland	1855-74	Liczbińska 2010
DE/urban West 1846	**242**	Unwght average from IMRs from 4 provinces covered (Munster, Dusseldorf, Schwarzwaldkreis, Neckarkreis)	1875-77	Knodel 1974, 164, 289
DE/Vechta 1700	**186**	The parish of Hartum, some 80 kms from the border of the area	1700-49	Kloke 1998, p. 87
Denmark/Aalborg 1787	**157**	exact province	1836-40	Lokke 1998, Bilag 2.3a
Denmark/Aarhus 1787	**152**	exact province	1836-40	Ibid.
Denmark/Bornholm 1787			**n/a**	
Denmark/Frederiksborg 1787	**130**	exact, rural district	1836-40	Lokke 2002, map 1
Denmark/Hjørring 1787	**130**	exact county	1836-40	Ibid.
Denmark/Holbæk 1787			**n/a**	
Denmark/København 1787	**195**	approximation, county	1836-40, rural	Lokke 2002, map 1 + Lokke 1998, Bilag 2.1
Denmark/Maribo 1787	**160**	exact county	1836-40, rural	Lokke 2002, map 1
Denmark/Odense 1787	**130**	exact county	1836-40, rural	Ibid.
Denmark/Præstø 1787	**130**	exact county	1836-40, rural	Ibid.
Denmark/Randers 1787	**160**	exact province	1850-54	Lokke 1998, Bilag 2.4a
Denmark/Ribe 1787	**129**	Unwght average from 10 districts of the county	1836-40, rural	Ibid., Bilag 2.3a
Denmark/Ringkøbing 1787	**170**	estimate, county, rural	1836-40, rural	Lokke 2002, map 1
Denmark/Roskilde 1787	**110**	approx. county	1836-40, rural	Ibid.
Denmark/Skanderborg 1787			**n/a**	
Denmark/Sorø 1787	**160**	approx. county	1836-40, rural	Lokke 2002, map 1

Denmark/Svendborg 1787	**n/a**			
Denmark/Thisted 1787	**140**	exact county	1850-54	Lokke 1998, Bilag 2.4a
Denmark/Tønder 1787	**n/a**			
Denmark/Vejle 1787	**130**	approx. county	1836-40, rural	Lokke 2002, map 1
Denmark/Viborg 1787	**159**	exact province	1836-40, rural	Lokke 1998, Bilag 2.3a
England/Bedfordshire 1851	**162**	exact county	1851-1861	Registrar General's Decennial Supplement for 1861
England/Berkshire 1851	**130**	exact county	1851-1861	Ibid.
England/Buckinghamshire 1851	**144**	exact county	1851-1861	Ibid.
England/Cambridgeshire 1851	**163**	exact county	1851-1861	Ibid.
England/Cheshire 1851	**163**	exact county	1851-1861	Ibid.
England/Cornwall 1851	**136**	exact county	1851-1861	Ibid.
England/Cumberland 1851	**134**	exact county	1851-1861	Ibid.
England/Derbyshire 1851	**147**	exact county	1851-1861	Ibid.
England/Devonshire 1851	**123**	exact county	1851-1861	Ibid.
England/Dorset 1851	**124**	exact county	1851-1861	Ibid.
England/Durham 1851	**156**	exact county	1851-1861	Ibid.
England/Essex 1851	**137**	exact county	1851-1861	Ibid.
England/Gloucestershire 1851	**142**	exact county	1851-1861	Ibid.
England/Hampshire 1851	**125**	exact county	1851-1861	Ibid.
England/Herefordshire 1851	**132**	exact county	1851-1861	Ibid.
England/Hertfordshire 1851	**134**	exact county	1851-1861	Ibid.
England/Huntingdonshire 1851	**152**	exact county	1851-1861	Ibid.
England/Kent (extra London) 1851	**133**	exact county	1851-1861	Ibid.
England/Lancashire 1851	**186**	exact county	1851-1861	Ibid.

England/Leicestershire 1851	**168**	exact county	1851-1861	Ibid.
England/Lincolnshire 1851	**152**	exact county	1851-1861	Ibid.
England/London (Parts Of Middlesex, Surrey and Kent) 1851	**155**	exact county	1851-1861	Ibid.
England/Middlesex (extra London) 1851	**136**	exact county	1851-1861	Ibid.
England/Norfolk 1851	**169**	exact county	1851-1861	Ibid.
England/Northamptonshire 1851	**160**	exact county	1851-1861	Ibid.
England/Northumberland 1851	**144**	exact county	1851-1861	Ibid.
England/Nottinghamshire 1851	**171**	exact county	1851-1861	Ibid.
England/Oxford 1851	**140**	exact county	1851-1861	Ibid.
England/Rutland 1851	**123**	exact county	1851-1861	Ibid.
England/Shropshire 1851	**128**	exact county	1851-1861	Ibid.
England/Somerset 1851	**128**	exact county	1851-1861	Ibid.
England/Staffordshire 1851	**174**	exact county	1851-1861	Ibid.
England/Suffolk 1851	**144**	exact county	1851-1861	Ibid.
England/Surrey (extra London) 1851	**121**	exact county	1851-1861	Ibid.
England/Sussex 1851	**124**	exact county	1851-1861	Ibid.
England/Warwick 1851	**167**	exact county	1851-1861	Ibid.
England/Westmorland 1851	**99**	exact county	1851-1861	Ibid.
England/Wiltshire 1851	**126**	exact county	1851-1861	Ibid.
England/Worcestershire 1851	**148**	exact county	1851-1861	Ibid.
England/Yorkshire East Riding 1851	**176**	exact county	1851-1861	Ibid.
England/Yorkshire North	**128**	exact county	1851-1861	Ibid.

Riding 1851				
England/Yorkshire West Riding 1851	**170**	exact county	1851-1861	Ibid.
ES/Andalucía 1752	**287**	Andalusia province	1860	Pradas 2005
ES/Barcelona province urban 1889	**183**	Exact	1901	Arbelo Curbelo 1966, p. 164,
ES/Catalonia rural 1880-1890	**110**	exact, province Barcelona	1901	Ibid.
ES/other Catalonia urban 1889	**146**	Unwght average of Gerona, Lerida, and Terragona provinces without capitals	1901	Ibid.
F/Northeast 1846	**233**	Unwght average from 9 departments where villages were located	1806-1810, except for Pas-de-Calais 1851-72	Bourdelais, P., Garden, M. & Bideau, A. 1988, vol. 5, p. 281.
F/Northwest 1846	**202**	Unwght average from 8 departments where villages were located	1806-1810, except for Pas-de-Calais 1851-72	Ibid., p. 229-286.
F/South 1846	**155**	Unwght average from 8 departments where villages were located	1806-11	Dinet 1973; Fortin & Gillet 1967; Sangoi 2005.
F/Southwest 1831-1901	**120**	Pyrenees Atlantique	1901-05	Rollet 1978
F/St. Emilion (city) 1846/1856	**117**	city	1831-32	Pontet 1975, pp. 369-397.
GR/Kythira 1724	**198**	Greece	1860-64	Valaoras 1964, 132.
H/Great Plain 1869	**289**	region of Duna–Tisza köze	1892	Farago 2003, p . 21, Tab 5.
H/North-East 1869	**233**	South. part of Zemplen district	1900-10	Ferenc 2013
H/Northern Transdanubia 1869	**272**	Region of Duna Bal Partja	1892	Farago 2003, p . 21, Tab 5.
H/Southern Transdanubia 1869	**297**	Region of Duna Jobb Partja	1892	Ibid.

HR/Dubrovnik area 1674	**236**	Croatia country average	1820-1880	Todorova 2006, 82
Iceland 1703	**300**	various estimates	1771-75	Gardarsdóttir 2002, 55 ff.
Islands in the British Seas/Guernsey 1851	**n/a**			
Islands in the British Seas/Isle of Man 1851	**n/a**			
Islands in the British Seas/Jersey 1851	**n/a**			
LT/Kovno 1847	**155**	Kovno province	1867-81	Rashin 1956, tab 151
LT/Vilna 1847	**125**	exact province	1867-81	Ibid.
LV/Courland Goldingen+Pilten 1797	**166**	exact province	1867-81	Ibid.
LV/Courland Mitau 1797	**166**	exact province	1867-81	Ibid.
LV/Courland Selburg 1797	**166**	exact province	1867-81	Ibid.
LV/Courland Tuckum 1797	**166**	exact province	1867-81	Ibid.
NL/Eindhoven and Helmond (city) 1811	**207**	Unwght average from the two exact municipalities	1841-60	Ekamper & Van Poppel 2008
NL/Goes (city) 1810	**271**	Goes municipality	1841-60	Ibid.
NL/North Brabant 1810	**168**	exact province	1841-60	Walhout 2019
NL/Tilburg (city) 1811	**118**	Tilburg municipality	1841-60	Ekamper & Van Poppel 2008
NL/Zeeland 1811	**234**	exact province	1811-19	Van Poppel, Jonker, and Mandemaakers 2005
Norway/Aggershuus 1801	**208**	Unwght average from IMRs of 4 parishes in the province	1790-1809	Thorvaldsen 2002, Appendix 3.
Norway/Bratsberg 1801	**185**	southern Norway	1802-03	Ibid., Tab. 1.
Norway/Buskerud 1801	**194**	Unwght average from 4 parishes from the province	1791-1801	Thorvaldsen 2002, Appendix 3.
Norway/Christiania 1801	**242**	1 parish from the county	1791-1800	Ibid.

Norway/Christians 1801	**160**	Unwght average from 4 parishes from the province	1799-1803	Ibid.
Norway/Finmarken 1801	**176**	Unwght average from 4 parishes from the province	1791-1800	Ibid.
Norway/Hedemarken 1801	**140**	Rendalen parish from the province	1733-1780	Sogner 1984
Norway/Jarlsberg O 1801	**158**	Unwght average from 5 parishes from the province	1791-1803	Thorvaldsen 2002, Appendix 3.
Norway/Lister Og M 1801	**190**	1 parish from the province	1790-1800	Ibid.
Norway/Nedenæs 1801	**185**	south Norwegian parishes cluster	1802-1803	Ibid., tab 1
Norway/Nordland 1801	**114**	Unwght average from 2 parishes in the province	1799-1803	Ibid., Appendix 3.
Norway/Nordre Berg 1801	**177**	Unwght average from 10 parishes from the province	1799-1803	Ibid.
Norway/Nordre Tron 1801	**145**	1 parish in the province	1790-99	Ibid.
Norway/Romsdal 1801	**212**	Unwght average from 3 parishes from the province	1791-1803	Ibid.
Norway/Smaalehnene 1801	**220**	1 parish from the province	1800-25	Ibid.
Norway/Søndre Berg 1801	**180**	Unwght average from 8 parishes from the province	max 1770-1804	Ibid.
Norway/Søndre Tron 1801	**191**	Unwght average from 3 parishes from the province	1799-1803	Ibid.

Norway/Stavanger 1801	**193**	Unwght average from 2 parishes in the province	1799-1803	Ibid.
Norway/Trondheim 1801	**198**	Unwght average from 3 parishes from the province	1799-1803	Ibid.
PL/Central Belarus 1768-1804	**342**	the Koren parish, central Belarus	1762-91	Nosevich 2004, p. 140
PL/Chelmska Land 1791-1792	**342**	the Koren parish, central Belarus	1762-91	Ibid.
PL/Greater Poland 1666-1809	**241**	Greater Poland (two parishes)	1855-74	Liczbińska 2010
PL/Kujavia 1766-1792	**250**	Poznan province, various parishes	after 1855	Liczbińska 2010, 2017
PL/Lesser Poland 1789-1792	**250**	parish of Bejsce, 94 kms from the region's centroid	late 18th c.	Piasecki 1990
PL/Ostrzeszow County 1790-1791	**250**	parish of Bejsce, 300 kms from the region's centroid	late 18th c.	Ibid.
PL/Podolia 1785-1819	**169**	province	1867-81	Rashin 1956, tab 151
PL/Polesia 1795	**342**	the Koren parish, central Belarus	1762-91	Nosevich 2004, p. 140
PL/Silesian lowlands 1792	**210**	Bujakow parish, 76 kms from the region's centroid	1743-1799	Szołtysek, personal comm.
PL/Sudenten 1805	**278**	Liegnitz province	1862-66	Knodel 1974, 164, 289
PL/Sudeten 1781-82			**n/a**	
PL/Warmia 1695-1772	**230**	parish Dobre Miasto, 17-18th c. (1 of the 2 in the region)	1695	Borowski 1975
PL/Wielunskie County 1790-1792	**250**	parish of Bejsce, 207 kms from the region's	late 18th c.	Piasecki 1990

		centroid		
PL/Zhytomyr County 1791	**237**	Hetmanate, one of the parishes in the centre of the region	1780	Y. Voloshyn (unpubl. paper)
RO/Eastern Wallachia 1838	**200**	4 districts corresponding to the sample locations, rural	1891	Mișcarea Populațiunei României 1895, calculations by B. Matescu
RO/Moldavia Catholics 1781-1787	**291**	Roman district, rural	1891	Ibid.
RO/Moldavia Catholics 1866-1879	**291**	Roman district, rural	1891	Ibid.
RO/Northern Wallachia 1838	**180**	2 districts corresponding to the sample locations, rural	1891	Ibid.
RO/Partium 1869	**233**	Szilagy county	1900-10	Ferenc 2013
RO/Southern Wallachia 1838	**210**	4 districts corresponding to the sample locations, rural settlements only	1891	Mișcarea Populațiunei României 1895, calculations by B. Matescu
RO/Southwestern Wallachia 1838	**180**	2 districts corresponding to the sample locations, rural settlements only	1891	Ibid.
RO/Transylvania 1869	**248**	Region of Transylvania (Erdely)	1892	Farago 2003, p . 21, Tab 5
RUS/Braclav Governorate 1795	**169**	Podolia governorate	1867-81	Rashin 1956, tab 151
RUS/Cherdyn (city) 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Chusovoy 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Gagarin villages 1814	**334**	District of Moscow	1851-58	Blum & Troitskaya 1997
RUS/Ilyinsky 1710	**438**	Perm' province	1867-81	Rashin 1956, tab 151
RUS/Ilyinsky West 1710	**438**	Perm' province	1867-81	Ibid.

RUS/Kama West 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Kaygorodok 1710	**334**	District of Moscow	1745-63	Blum & Troitskaya 1997
RUS/Moscow area 1897	**366**	Moscow guberniya	1886-97	Rashin 1956, tab 151
RUS/Solikamsk (city) 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Solikamsk Center 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Solikamsk North 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Solikamsk South 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Southern Perm 1710	**438**	Perm' province	1896-97	Ibid.
RUS/Southwest Perm 1710	**438**	Perm' province	1867-81	Ibid.
RUS/St. Petersburg area South 1811	**345**	Petersburg province	1867-81	Ibid.
RUS/St. Petersburg area West 1811	**345**	Petersburg province	1867-81	Ibid.
RUS/Tula (city) 1710	**302**	Tulska province	1867-81	Ibid.
RUS/Ural Centre 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Ural East 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Ural iron plants 1710	**438**	Perm' province	1896-97	Ibid.
RUS/Ural North 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Ural North Centre 1710	**438**	Perm' province	1896-97	Ibid.
RUS/Ural South 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Ural Verkhoturye (city) 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Ural West 1710	**438**	Perm' province	1867-81	Ibid.
RUS/Vyatka (city) 1710	**383**	Vyatka governorate	1867-81	Ibid.
RUS/West Siberia 1710	**274**	6 parishes of Tobols province	1816	Кабакова 2008, p. 115, tab 29
RUS/Yarensk East 1710	**438**	Perm' province	1867-81	Rashin 1956, tab 151
RUS/Yarensk West 1710	**438**	Perm' province	1867-81	Ibid.
Scotland/Aberdeenshire	**111**	exact county	1880	Histpop

1881				
Scotland/Angus and Forfarshire 1881	**150**	exact county	1880	Ibid.
Scotland/Argyll 1881	**82,5**	exact county	1880	Ibid.
Scotland/Ayr 1881	**119**	exact county	1880	Ibid.
Scotland/Banff 1881	**120**	exact county	1880	Ibid.
Scotland/Berwickshire 1881	**109**	exact county	1880	Ibid.
Scotland/Bute 1881	**93**	exact county	1880	Ibid.
Scotland/Caithness 1881	**79**	exact county	1880	Ibid.
Scotland/Clackmannanshire 1881	**113**	exact county	1880	Ibid.
Scotland/Dumfriesshire 1881	**103**	exact county	1880	Ibid.
Scotland/Dunbartonshire 1881	**119**	exact county	1880	Ibid.
Scotland/East Lothian and Haddingtonshire 1881	**113**	exact county (East Lothian (Haddington))	1880	Ibid.
Scotland/Fife 1881	**116**	exact county	1880	Ibid.
Scotland/Inverness-shire 1881	**102**	exact county	1880	Ibid.
Scotland/Kincardinshire 1881	**92**	exact county	1880	Ibid.
Scotland/Kinross-shire 1881	**93**	exact county	1880	Ibid.
Scotland/Kirkudbrightshire 1881	**86**	exact county	1880	Ibid.
Scotland/Lanarkshire 1881	**139**	exact county	1880	Ibid.
Scotland/Mid Lothian and Edinburghshire 1881	**139**	exact county	1880	Ibid.
Scotland/Morayshire and Elginshire 1881	**125**	exact county	1880	Ibid.
Scotland/Nairnshire 1881	**99**	exact county	1880	Ibid.

Scotland/Orkney 1881	**62**	exact county	1880	Ibid.
Scotland/Peeblesshire 1881	**76**	exact county	1880	Ibid.
Scotland/Perthshire 1881	**111**	exact county	1880	Ibid.
Scotland/Renfrewshire 1881	**138**	exact county	1880	Ibid.
Scotland/Ross and Cromarty 1881	**91**	exact county	1880	Ibid.
Scotland/Roxburghshire 1881	**102**	exact county	1880	Ibid.
Scotland/Selkirk 1881	**125**	exact county	1880	Ibid.
Scotland/Shetland 1881	**64**	exact county	1880	Ibid.
Scotland/Stirlingshire 1881	**113**	exact county	1880	Ibid.
Scotland/Sutherland 1881	**87**	exact county	1880	Ibid.
Scotland/West Lothian and Linlithgowshire 1881	**111**	exact county	1880	Ibid.
Scotland/Wigtown 1881	**93**	exact county	1880	Ibid.
SE/Belgrade 1733	**246**	Belgrade (city)	1880-1887	Državopis Srbije 17, 626-649 and 671
SE/Jasenica 1863	**157**	Kragujevački okrug (incl. village location)	1880-1887	Ibid.
SE/Jasenica 1884	**157**	Kragujevački okrug (incl. village location)	1880-1887	Ibid.
SE/Kruševac (city) 1863	**161**	Kruševački okrug	1880-1887	Ibid.
SK/Central 1869	**177**	approx. county	1900-10	Ferenc 2013
SK/East 1869	**190**	exact, average from 4 counties of Hungarian Slovakia (Zemplin, Spis, Saris, Abauj[Abov])		Veres 1985
SK/West 1869	**210**	exact, average from 4 counties of Hungarian Slovakia (Bratislava, Nitra, Komarno, Tekov)	1900-04	Ibid.

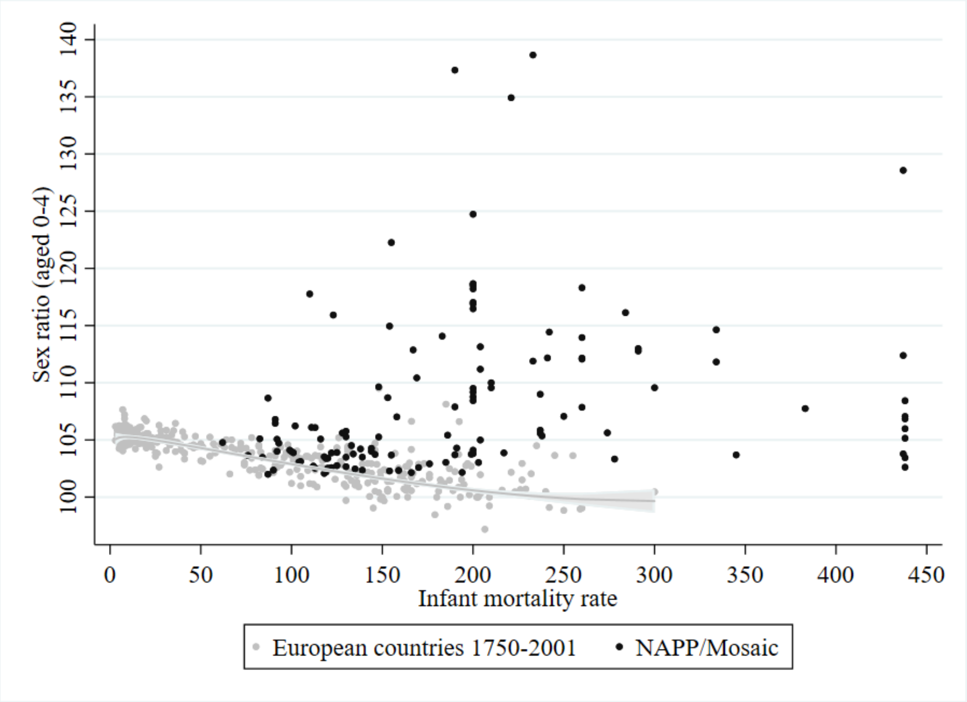
Sweden/Älvsborg 1880	**100**	exact province	1881	von Hofsten, Erland & Lundström, Hans, 1976
Sweden/Blekinge 1880	**133**	exact province	1881	Ibid.
Sweden/Gävleborg 1880	**122**	exact province	1881	Ibid.
Sweden/Göteborg och Bohus 1880	**112**	exact province	1881	Ibid.
Sweden/Gotland 1880	**84**	exact province	1881	Ibid.
Sweden/Halland 1880	**117**	exact province	1881	Ibid.
Sweden/Jämtland 1880	**91**	exact province	1881	Ibid.
Sweden/Jönköping 1880	**83**	exact province	1881	Ibid.
Sweden/Kalmar 1880	**92**	exact province	1881	Ibid.
Sweden/Kopparberg 1880	**122**	exact province	1881	Ibid.
Sweden/Kristianstad 1880	**101**	exact province	1881	Ibid.
Sweden/Kronoberg 1880	**105**	exact province	1881	Ibid.
Sweden/Malmöhus 1880	**120**	exact province	1881	Ibid.
Sweden/Norrbotten 1880	**118**	exact province	1881	Ibid.
Sweden/Örebro 1880	**104**	exact province	1881	Ibid.
Sweden/Östergötland 1880	**87**	exact province	1881	Ibid.
Sweden/Skaraborg 1880	**93**	exact province	1881	Ibid.
Sweden/Södermanland 1880	**101**	exact province	1881	Ibid.
Sweden/Stockholm 1880	**169**	unweighted average from exact province and city of Stockholm	1881	Ibid.
Sweden/Uppsala 1880	**120**	exact province	1881	Ibid.
Sweden/Värmland 1880	**90**	exact province	1881	Ibid.
Sweden/Västerbotten 1880	**119**	exact province	1881	Ibid.
Sweden/Västernorrland 1880	**146**	exact province	1881	Ibid.
Sweden/Västmanland 1880	**128**	exact province	1881	Ibid.
TR/Istanbul (city) 1885	**217**	Turkey	1955	https://www.statista.com

TR/Istanbul (city) 1907	**n/a**			
UKR/Berdychiv (city) 1897	**204**	Kiew governorate	1867-81	Rashin 1956, tab 151
UKR/Berdychiv region North rural 1897	**204**	Kiew governorate	1867-81	Ibid.
UKR/Berdychiv region North urban 1897	**204**	Kiew governorate	1867-81	Ibid.
UKR/Berdychiv region South rural 1897	**204**	Kiew governorate	1867-81	Ibid.
UKR/Berdychiv region South urban 1897	**204**	Kiew governorate	1867-81	Ibid.
UKR/Hetmanate North 1765	**237**	Hetmanate, Пирятинська протопопіа (one of the parishes in the centre of the region)	1780	Y. Voloshyn (unpubl. paper)
UKR/Hetmanate Southeast 1765	**237**	Hetmanate, Пирятинська протопопіа (one of the parishes in the centre of the region)	1780	Ibid.
UKR/Hetmanate Southwest 1765	**237**	Hetmanate, Пирятинська протопопіа (one of the parishes in the centre of the region)	1780	Ibid.
UKR/Hetmanate West 1765	**237**	Hetmanate, Пирятинська протопопіа (one of the parishes in the centre of the region)	1780	Ibid.
USSR/Central 1926/27	**260**	Сибирский край	1926	Lorimer, p. 81, plate XII
USSR/East 1926/27	**260**	Сибирский край	1926	Ibid.

USSR/North 1926/27	**260**	Сибирский край	1926	Ibid.
USSR/Northeast 1926/27	**260**	Сибирский край	1926	Ibid.
USSR/Northwest 1926/27	**203**	Сибирский край	1926	Ibid.
USSR/South 1926/27	**260**	Сибирский край	1926	Ibid.
USSR/Southeast 1926/27	**260**	Сибирский край	1926	Ibid.
USSR/West 1926/27	**300**	Сибирский край	1926	Ibid.
Wales/Anglesey 1851	**111**	exact county	1851-1861	Registrar General’s Decennial Supplement for 1861
Wales/Brecknockshire 1851	**148**	exact county	1851-1861	Ibid.
Wales/Cardiganshire 1851	**88**	exact county	1851-1861	Ibid.
Wales/Carmarthenshire 1851	**108**	exact county	1851-1861	Ibid.
Wales/Carnarvonshire 1851	**129**	exact county	1851-1861	Ibid.
Wales/Denbighshire 1851	**138**	exact county	1851-1861	Ibid.
Wales/Flintshire 1851	**135**	exact county	1851-1861	Ibid.
Wales/Glamorganshire 1851	**154**	exact county	1851-1861	Ibid.
Wales/Merionethshire 1851	**104**	exact county	1851-1861	Ibid.
Wales/Monmouthshire 1851	**144**	exact county	1851-1861	Ibid.
Wales/Montgomeryshire 1851	**118**	exact county	1851-1861	Ibid.
Wales/Pembrokeshire 1851	**121**	exact county	1851-1861	Ibid.
Wales/Radnorshire 1851	**111**	exact county	1851-1861	Ibid.

Note: the following people are acknowledged for their help in collecting the above mentioned data: E. Garrett, S. Edvisson, W. Letsch, I. Devos, P. Ori, L. Kesztenbaum, G. Thorvaldsen, E. Walhout, A. Løkke, I. Gregory.

FIGURE 2C: National trends in infant mortality and the NAPP/Mosaic extreme sex ratios



Note: The fitting line is based on a fractional polynomial regression. For convenience, only NAPP/Mosaic data points exhibiting relatively high sex ratios are presented, using the threshold of 102.
Source: Beltrán Tapia (2019) and Szołtysek (2020), NAPP/Mosaic Historical Infant Moratlity Datset.

D. Robustness tests

FIGURE D1: Residuals from the Spatial Model

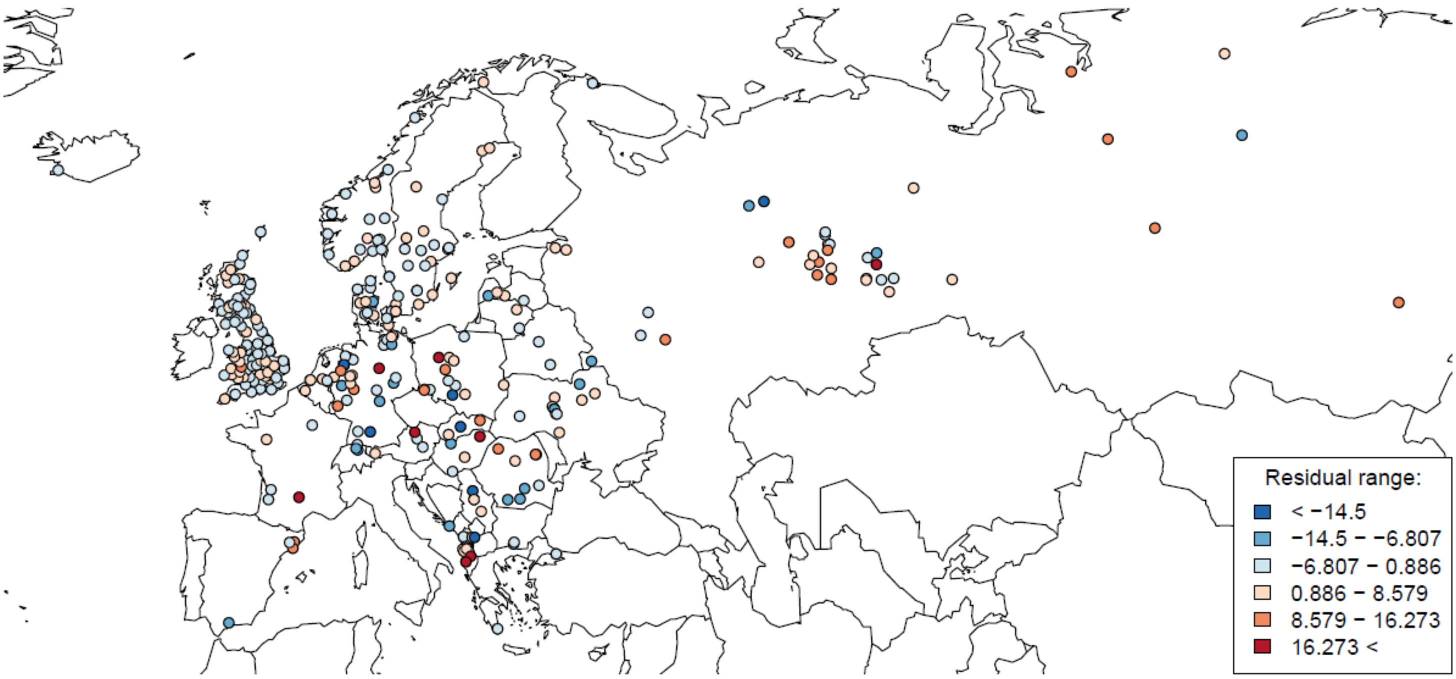


TABLE D2. Regression results excluding NAPP (different age-groups: 0-1, 1-5 and 5-9)

	CSR 0			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,1902	0,1226	0,1207	
Infant Mortality Rate	0,00010	0,0002	0,6666	
Other censuses	-0,0876	0,0498	0,0783	.
Pre-modern censuses	-0,0274	0,0647	0,6713	
Infants (over children aged 0-4)	0,0444	0,3041	0,8838	
Children aged 0-10 (over pop. aged 15-64)	0,2374	0,1845	0,1982	
Ruggedness	0,0212	0,0157	0,1771	
F/M Age-heaping (Wtot)	0,0651	0,0540	0,2281	
D2	0,0940			
Moran's I	-0,0030			
n	132			

	CSR 1-5			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,02077	0,03431	0,54490	
Infant Mortality Rate	0,00005	0,00007	0,48913	
Other censuses	-0,06411	0,01625	0,00008	***
Pre-modern censuses	-0,07457	0,01815	0,00004	***
Infants (over children aged 0-4)	-0,05138	0,07220	0,47668	
Children aged 0-10 (over pop. aged 15-64)	0,02480	0,04992	0,61935	
Ruggedness	0,00524	0,00502	0,29657	
F/M Age-heaping (Wtot)	0,0526	1,61E-02	0,00108	**
D2	0,230			
Moran's I	0.108 *			
n	160			

	CSR 5-9			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,0634	0,0356	0,0747	.
Infant Mortality Rate	0,0000	0,0001	0,7087	
Other censuses	-0,0456	0,0171	0,0078	**
Pre-modern censuses	-0,0246	0,0195	0,2068	
Infants (over children aged 0-4)	-0,0286	0,0656	0,6629	
Children aged 0-10 (over pop. aged 15-64)	0,1711	0,0534	0,0014	**
Ruggedness	0,0102	0,0052	0,0507	.
F/M Age-heaping (Wtot)	0,0323	0,0170	0,0578	.
D2	0,1650			
Moran's I	0.149 **			
n	160			

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

FIGURE D2. Residuals from the regression model excluding NAPP

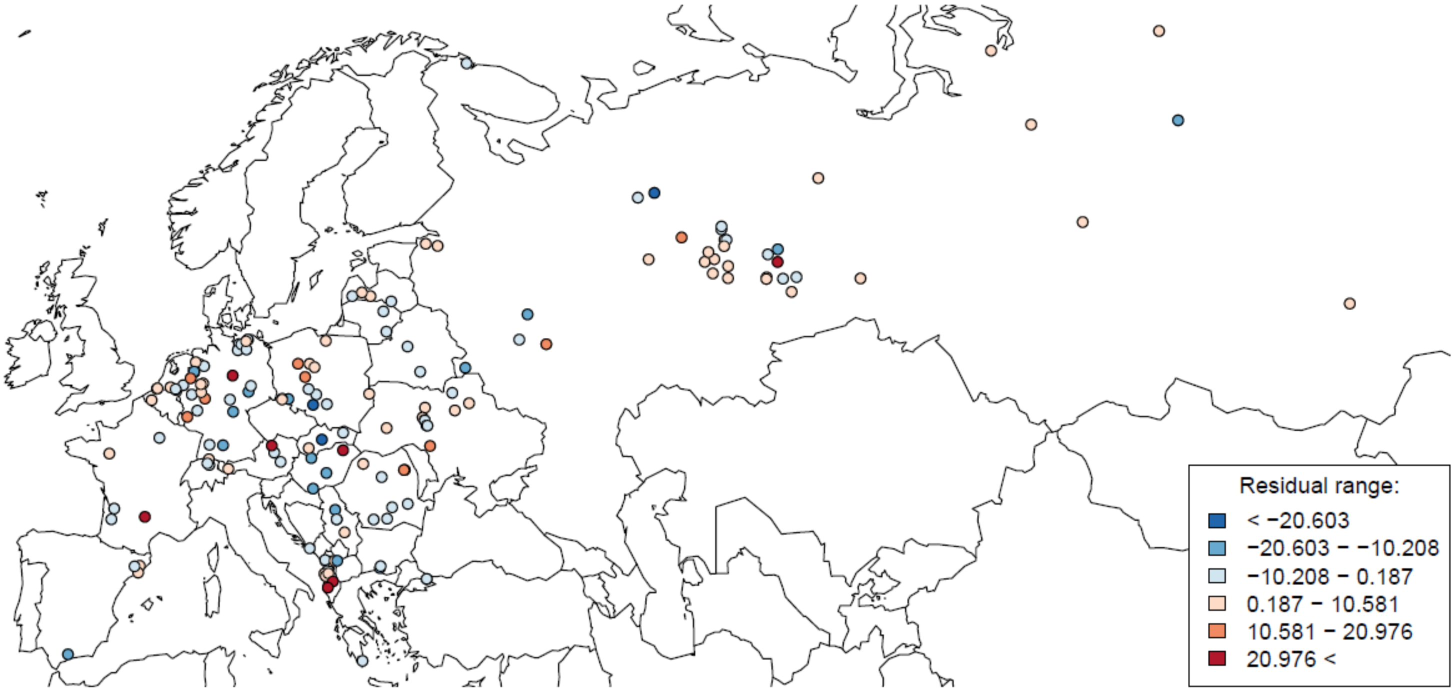


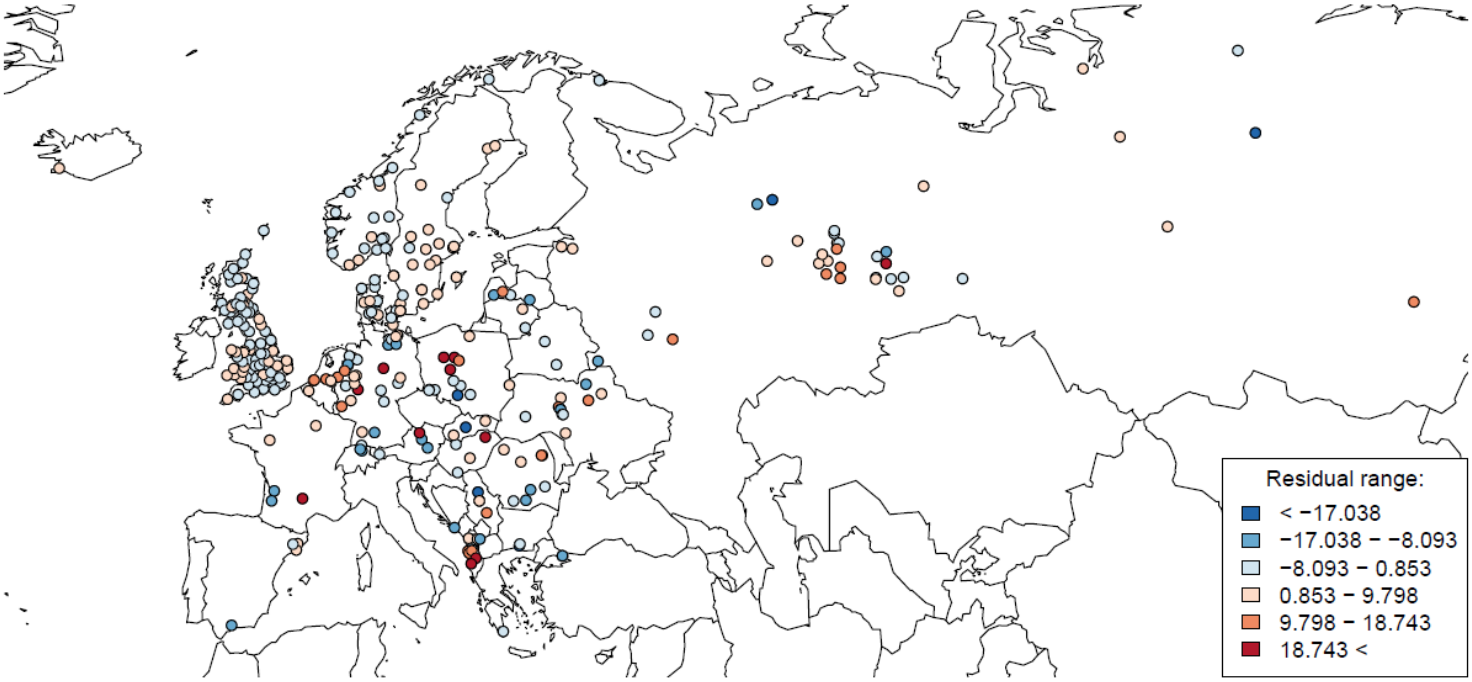
TABLE D3. Regression results including the Patriarchy Index (different age-groups: 0-1, 1-5 and 5-9)

	SR 0			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,0041	0,0438	0,9247	
Infant Mortality Rate	-0,00038	0,0001	0,0000	***
Other censuses	-0,0478	0,0203	0,0184	*
Pre-modern censuses	-0,0855	0,0139	0,0000	***
Infants (over children aged 0-4)	0,3076	0,1293	0,0174	*
Children aged 0-10 (over pop. aged 15-64)	-0,0086	0,0638	0,893	
Ruggedness	0,0158	0,0034	0,0000	***
F/M Age-heaping (Wtot)	-0,0638	0,018	0,0004	***
Patriarchy Index	0,0075	0,0015	0,0000	***
D2	0,3			
Moran's I	0,018			
n	280			

	SR 1-5			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,01397	0,01599	0,382	
Infant Mortality Rate	-0,00039	0,00003	0,0000	***
Other censuses	-0,03783	0,00754	0,0000	***
Pre-modern censuses	-0,08643	0,00448	0,0000	***
Infants (over children aged 0-4)	0,24029	0,03421	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	-0,02606	0,02435	0,285	
Ruggedness	0,01958	0,00149	0,0000	***
F/M Age-heaping (Wtot)	-0,063	0,0000	0,0000	***
Patriarchy Index	0,0085	0,0000	0,0000	***
D2		0,56		
Moran's I		0.161 ***		
n		308		
	SR 5-9			
	Estimate	Std. Error	Pr(>|z|)	
(Intercept)	-0,0587	0,0169	0,0005	***
Infant Mortality Rate	-0,0004	0,0000	0,0000	***
Other censuses	-0,0304	0,0078	0,0001	***
Pre-modern censuses	-0,0771	0,0048	0,0000	***
Infants (over children aged 0-4)	0,199	0,0356	0,0000	***
Children aged 0-10 (over pop. aged 15-64)	0,0552	0,0258	0,0323	*
Ruggedness	0,0198	0,0016	0,0000	***
F/M Age-heaping (Wtot)	-0,0612	0,0079	0,0000	***
Patriarchy Index	0,0092	0,0006	0,0000	***
D2		0,542		
Moran's I		0.225 ***		
n		308		

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

FIGURE D3. Residuals from the regression model including the Patriarchy Index



D4. COMPONENTS OF THE PATRIARCHY INDEX

Domain	Component	Abbreviation	Definition/measurement	Relationship with patriarchy	Specification
Male domination	Proportion of female household heads	*Female heads*	The proportion of all female household heads (20+ years) among all adult heads of family households	Negative	Age-standardized
	Proportion of young brides	*Young brides*	The proportion of ever-married women in the age group 15-19 years	Positive	
	Proportion of wives who are older than their husbands	*Older wives*	The proportion of all wives who are older than their husbands among all couples for whom the ages of both spouses are known	Negative	Age-standardized
	Proportion of young women living as non-kin	*Female non-kin*	The proportion of women aged 20-34 years who live as non-kin, usually as lodgers or servants	Negative	Age-standardized
Generational domination	Proportion of elderly men co-residing with a younger household head	*Younger household head*	The proportion of men aged 65+ years living in a household headed by a male household head of a younger generation	Negative	Only family households; the elderly men must be relatives of the household head
	Proportion of neolocal residence among young men	*Neolocal*	The proportion of male household heads living without any relatives except spouses/children among ever-married men aged 20-29 years	Negative	Only family households; age-standardized
	Proportion of elderly people living with lateral relatives	*Lateral*	The proportion of people aged 65+ years living with at least one lateral relative in the household	Positive	Only family households
Patrilocality	Proportion of elderly people living with married daughters	*Married daughter*	The proportion of people aged 65+ years living with at least one married daughter in the same household among those elderly people who live with at least one married child in the same household	Negative	Only family households
Son preference	Proportion of boys among the last child	*Boy as last child*	The proportion of boys among the last children (if the last child is one of a set of siblings of both sexes, he or she will be excluded from the analysis).	Positive	Only children (aged 10-14 years) of household heads; family households

Source: Szołtysek et. al. 2017.

In text references

Buławski, R. 1930. Projekt drugiego powszechnego spisu powszechnego na tle doświadczeń spis 1921 r. oraz praktyki zagranicznej. *Kwartalnik Statystyczny* 7: 17–151.
Coale, A. J., and S. C. Watkins (eds.). 1986. *The Decline of Fertility in Europe*. Princeton: Princeton University Press.
Coale et al., 1979. *Human Fertility in Russia Since the Nineteenth Century*. Princeton Univ. Press, Princeton, N. J
Corsini, C.A., Viazzo, P.P. (eds.). 1997. *The Decline of Infant and Child Mortality: The European Experience, 1750-1990*. The Hague: Martinus Nijhoff Publishers.
Demeny, P., Shorter, F. C. 1968. *Estimating Turkish Mortality, Fertility, and Age Structure*. Istanbul: Istanbul University Press.
Ewbank, D. C. 1981. *Age misreporting and age-selective underenumeration: sources, patterns, and consequences for demographic analysis*. Washington, D.C.: National Academy Press.
Falkingham, J., Gjonça, A. 2001. Fertility transition in Communist Albania, 1950-90. Population Studies 55(3): 309–318.
Johansson, S., Nygren, O. 1991. The Missing Girls of China: A New Demographic Account. Population and Development Review 17(1): 35-51.
Kluesener, S. and multiple authors. 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): 1849-1864.
Miller, B.D. 1989. Changing Patterns of Juvenile Sex Ratios in Rural India: 1961 to 1971. Economic and Political Weekly 24 (22): 1229-1236.
Spoorenberg, T. 2016. On the masculinization of population: the contribution of demographic development – a look at sex ratios in Sweden over 250 years. Demographic Research 34(37): 1053–1062.
Szołtysek, M., Klüsener, S., Poniat, R., Gruber, S. 2017. The Patriarchy Index: A New Measure of Gender and Generational Inequalities in the Past. Cross-Cultural Research 51(3):228-262.
Szołtysek, Poniat, R., Gruber, S. 2018. Age heaping patterns in Mosaic data. Historical Methods: A Journal of Quantitative and Interdisciplinary History 51(1): 13-38.

D5. DATA FULL REFERENCES

Mosaic data:

Siegfried Gruber, Karl Kaser, Gentiana Kera, Enriketa Pandelejmoni. *1918 Census of Albania, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2019.

Laboratory of Historical Demography (MPIDR). *1869 Census of Hungary, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1910 Census of Austria, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Familiekunde Vlaanderen and Laboratory of Historical Demography (MPIDR). *1814 Census of Western Flanders, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Ulf Brunnbauer. *Household registers of Rhodope region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1674* Status Animarum for Lisac and Pridvorje, *Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Laboratory of Historical Demography (MPIDR). *1846 Census of France, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

University of Bordeaux. *1831 Census of Sallespisse, Version 1.2* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2013.

University of Bordeaux. *1836 Census of Boulazac, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1841 Census of St. Jean de Luz, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1841 Census of Targon, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1876 Census of Boulazac, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1901 Census of Sauternes, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. 1846 Census of Saint-Émilion, Version 1.2 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

University of Bordeaux. 1856 Census of Saint-Émilion, Version 1.2 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1846 German Customs Union Census, Version 2.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1846 Census of Höhscheid, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1858 German Customs Union Census, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1861 Census of Haigerloch, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. *1819 Census of Mecklenburg-Schwerin, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2016.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. 1819 Census of Rostock, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. 1867 Census of Mecklenburg-Schwerin, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2016.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. 1867 Census of Rostock, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. *1900 Census of Rostock, Version 1.0* [Mosaic Historical Microdata File]. Rostock, Germany: www.censusmosaic.org, 2013.

Laboratory of Historical Demography (MPIDR). *1749 Status Animarum of Münster, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1690-1713 Status Animarum of Oldenburger Münsterland, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *Status Animarum for Oggelshausen, Dischingen, Gögglingen, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Elisabeth Frische. *1861 Census of Lippborg, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

Volker Wilmsen. *1864 Census of Albachten, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Violetta Hionidou. *1724 Census of Cerigo, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2020.

Laboratory of Historical Demography (MPIDR). 1811 Census of Zeeland, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Laboratory of Historical Demography (MPIDR). 1810 Census of North Brabant, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Mikołaj Szołtysek. *CEURFAMFORM database, Version 23* [SPSS file]. Rostock, 2012.

Laboratory of Historical Demography (MPIDR). 1781-1879 Status Animarum in Moldavia, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Laboratory of Historical Demography (MPIDR). *1838 Census of Wallachia, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1710 Russian household census of the Ural region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Mikołaj Szołtysek, Siegfried Gruber. *1710 Russian household census, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2020.

Laboratory of Historical Demography (MPIDR). *1765 Rumyantsev census of Hetmanate region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Laboratory of Historical Demography (MPIDR). *1795 Braclav Region Revision Lists, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1797 Revision lists of Courland, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Maria Markova, Siegfried Gruber, Mikołaj Szołtysek. *1811 St. Petersburg region confessional lists, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2020.

Laboratory of Historical Demography (MPIDR). *1814 Russian list of inhabitants, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). 1847 Lithuanian Estate Household Listings, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Laboratory of Historical Demography (MPIDR). *1897 Russian Empire Census, Moscow region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org,2014.

Tomasz Jankowski. *1897 Russian Empire Census, Berdychiv region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Siegfried Gruber. *1733/34 Orthodox Status Animarum of Belgrade, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2018.

Joel M. Halpern and Siegfried Gruber. *1863 Census of Jasenički srez and the city of Kruševac, Serbia, Version 1.2* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Joel M. Halpern and Siegfried Gruber. *1884 Census of Jasenički srez, Serbia, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

David Anderson, Siegfried Gruber, Mikołaj Szołtysek. *1926-27 Polar Census, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2018.

Jörg Baten, Franziska Tollnek. *1752 Catastro de Ensenada of Colomera and Estepona, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

Laboratory of Historical Demography (MPIDR). 1880-1890 Local Censuses in Catalonia, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Ulrich Pfister. *1634-1764 Soul listings of canton Zürich, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Vienna Database on European Family History. *1671-1685 Church listings of canton Zürich, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Vienna Database on European Family History. *1870 Census of Zürich, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Alan Duben. *1885 Census of Istanbul, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Alan Duben. *1907 Census of Istanbul, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

NAPP data:

Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D020.V7.2

The author wishes to acknowledge the statistical office that provided the underlying data making this research possible:

- Iceland 1703: Ólöf Garðarsdóttir (University of Iceland) and National Archives of Iceland (NAI). 1703 Census of Iceland, Version 1.0.

- Denmark 1787: Nanna Floor Clausen, Danish National Archives. 1787 Census of Denmark, Version 1.0. Minneapolis: Minnesota Population Center [distributor], 2015.

- Norway 1801: The Digital Archive (The National Archive), University of Bergen, and the Minnesota Population Center. Census of Norway 1801, Version 1.0. Bergen, Norway: University of Bergen, 2011.

- England and Wales 1851: K. Schürer and Higgs, E., (2014) Integrated Census Microdata (I-CeM), 1851-1911. [data collection]. UK Data Service. 2014. SN: 7481, http://doi.org/10.5255/UKDA-SN-7481-1

- Sweden 1880: The Swedish National Archives, Umeå University, and the Minnesota Population Center. National Sample of the 1880 Census of Sweden, Version 1.0. Minneapolis: Minnesota Population Center [distributor], 2014.

- Scotland 1881: K. Schürer and M. Woollard, National Sample from the 1881 Census of Great Britain [computer file], Colchester, Essex: History Data Service, UK Data Archive [distributor], 2003.