Science Journal of Public Health
2015; 3(2): 291-302
Published online March 18, 2015 (http://www.sciencepublishinggroup.com/j/sjph)
doi: 10.11648/j.sjph.20150302.30
ISSN: 2328-7942 (Print); ISSN: 2328-7950 (Online)

Glimpses at the History of Sex Ratio Studies

Johan Fellman

Hanken School of Economics, POB 479, FI-00101 Helsinki, Finland

Email address:
fellman@hanken.fi

To cite this article:
Johan Fellman. Glimpses at the History of Sex Ratio Studies. *Science Journal of Public Health*. Vol. 3, No. 2, 2015, pp. 291-302.
doi: 10.11648/j.sjph.20150302.30

Abstract: The sex ratio at birth (SR) is defined as the number of males per 100 females and is almost always around 106. John Graunt (1620–1674) was the first to compile data showing an excess of male births to female births and to note spatial and temporal variation in the SR. John Arbuthnot (1667–1735) demonstrated that the excess of males was statistically significant and asserted that the SR is uniform over time and space. Arbuthnot suggested that the regularity in the SR and the dominance of males over females must be an indication of divine providence. Nicholas Bernoulli's (1695–1726) counter-argument was that chance could give uniform dominance of males over females. Later, Daniel Bernoulli (1700–1782), Pierre Simon de Laplace (1749–1827) and Siméon-Denis Poisson (1781–1840) also contributed to this discussion. Attempts have been made to identify factors influencing the SR, but comparisons demand large data sets. Attempts to identify associations between SRs and stillbirth rates have failed to yield consistent results. A common pattern observed in different countries is that during the first half of the twentieth century the SR showed increasing trends, but during the second half the trend decreased. A common opinion is that secular increases are caused by improved socio-economic conditions. The recent downward trends have been attributed to new reproductive hazards. Factors that affect the SR within families remain poorly understood. Although they have an effect on family data, they have not been identified in large national birth registers.

Keywords: Still Birth Rate, Temporal Variation, Regional Variation, John Graunt, John Arbuthnot, Nicholas Bernoulli, Daniel Bernoulli, Pierre Simon de Laplace, Siméon-Denis Poisson

1. Introduction

The sex ratio at birth, also called the secondary sex ratio, and here denoted SR, is usually defined as the number of males per 100 females. Among newborns, there is almost always a slight excess of boys. Consequently, the SR is greater than 100, mainly around 106.

John Graunt [1] (1620-1674) was the first English vital statistician, and his book *Natural and Political Observations on the Bills of Mortality* was published in 1662. In addition, the complete collection of such material as existed up to 1668 plus all of the bills issued after that date were compiled by an unknown writer of the 18th century. Graunt´s work was of importance in that it gave the lead to other countries. For example, a Paris journal of 1666 stated that "the issue bills of mortality is a thing peculiar to the English", and the necessary alteration to enable a similar thing to be done in Paris was not added to the French legal code until 1667. The introduction into other European countries was after that date, sometimes much later ([2] p. 100).

John Arbuthnot [3] (1667-1735) demonstrated in his text *An Argument for Divine Providence, taken from the Constant Regularity observed in the Births of both Sexes* (1711) that the excess of males was statistically significant and asserted that the SR is uniform over time and space [4].

Arbuthnot stated that

> "AMONG innumerable Footsteps of Divine Providence to be found in the Works of Nature, there is a very remarkable one in the exact Ballance that is maintained between the Numbers of Men and Women; for by this means it is provided, that the Species may never fail, nor perish, since every Male may have its Female, and of a proportional Age. This Equality of Males and Females is not the Effect of chance but Divine Providence."

Although Arbuthnot considered the binomial distribution, he suggested that the regularity in the SR and the dominance of males over females could not be attributed to chance and must be an indication of divine providence. He stated that

> "But this Event is happily prevented by the wise Oeconomy of Nature; and to judge of the wisdom of the Contrivance,

> we must observe that the external Accidents to which Males are subject (who must seek their Food with danger) make a great havock of them, and that this loss exceeds far that of the other Sex occasioned by Diseases incident to it, as Experience convinces us. To repair that Loss, provident Nature, by the Disposal of its wise Creator, brings forth more Males than Females; and that in almost a constant proportion."

There are two main elements in Arbuthnot´s misapprehension. First is the invalid argument. He is simply ignorant of the fact that an event of chance, *p*, in sufficiently many repeated trials will very probably occur with a relative frequency very close to *p*. It requires no "art" to guarantee this consequence of Jacques Bernoulli´s limit theorem ([5] p. 168).

It is important to distinguish distinct intertwined questions. First, there is the question of whether a constant statistical stability can be the effect of chance. Following Arbuthnot, Derham (1657-1735) undoubtedly thought not. That was a mistake, and Nicholas Bernoulli (1695-1726), among others, said so at once. Second, there was the question of why the chance of a male should be about 14:13 (as Derham calculated) or 18:17 (as Bernoulli observed). Why a slight surplus of males over females? Hence, the fraction 18:17 is itself evidence of divine providence ([5] p. 170).

Nicholas Bernoulli's counter-argument was that Arbuthnot's model was too restrictive. Instead of a fair coin model, the model should be based on an asymmetric coin. Based on the generalized model, chance could give uniform dominance of males over females. Later, Daniel Bernoulli (1700-1782), Pierre Simon de Laplace (1749-1827) and Simeon-Denis Poisson (1781-1840) also contributed to this discussion [2, 5, 6].

In the 1870s, Berg [7] presented (in Swedish) in his study a detailed analysis of the sex ratio in Sweden. He considered the secondary sex ratio among births, but also in the living population. In the study of births, he considered all births, i.e. live births and stillbirths, and regional differences among the births. Although the study was published as late as 1871, he connected the SR to the blessing in the Holy Bible that said "Be fruitful and multiply, and fill the earth." According to Berg this mission should be protected by a law of Nature, including two sexes and the balance between them.

According to the literature, the sex ratio shows noticeable heterogeneity, and different scientists have suggested various influential factors.

2. Methods and Materials

Today, the study of secondary sex ratio is based on a more stable foundation of statistical theory. This enables scientists *i. a.* to identify influential factors.

Maximum likelihood estimation ([8]). If the theoretical proportion of males is p_0, then the observed relative frequency of males p is a maximum likelihood (ML) estimator of p_0 being unbiased, consistent, efficient and asymptotically normal with $E(p) = p_0$ and $Var(p) = \frac{p_0(1-p_0)}{N}$. According to the ML theory, $SR = \frac{p}{1-p}$ is a ML estimator of the transformed parameter $SR_0 = \frac{p_0}{1-p_0}$, but SR is not unbiased. Consider the difference

$$SR - SR_0 = \frac{p}{1-p} - \frac{p_0}{1-p_0} = \frac{p-p_0}{(1-p)(1-p_0)}.$$

Hence,. $E(SR) = SR_0 + E\left(\frac{p-p_0}{(1-p)(1-p_0)}\right) \neq SR_0$.

When $N \to \infty$, then $p \to p_0$ and $SR - SR_0 \to 0$ and the estimate SR is consistent, biased, but asymptotically unbiased, and normally distributed.

Standard deviations and confidence intervals ([8]). Visaria [9] stressed that random errors influence the variation in the SR. Therefore, he presented a numerical table of how the confidence intervals (CIs) of the SR depend on the observed SR and the number of births. He gave no formula for the intervals, but stated that "the standard error of an observed sex ratio can be estimated as the standard error of the proportion "*p*" of male births among the total". Fellman and Eriksson [8] interpreted Visaria´s statement such that he constructed CIs for *p*, that is

$$(p - k\sqrt{\frac{p(1-p)}{N}}, p + k\sqrt{\frac{p(1-p)}{N}}), \quad (1)$$

where k corresponds to the confidence level. After that, Visaria defined the CI for the SR, (SR_L, SR_U), so that $SR_L = \frac{p_L}{1-p_L}$ and $SR_U = \frac{p_U}{1-p_U}$, where $p_L = p - k\sqrt{\frac{p(1-p)}{N}}$ and $p_U = p + k\sqrt{\frac{p(1-p)}{N}}$.

Visaria´s attempt is based on the fact that SR is a monotonously increasing function of *p*. Numerical checking of his results confirmed this assumption.

Visaria´s CI has the following properties. If one introduces the following short notations $p_L = p - h$ and $p_U = p + h$, where $h = k\sqrt{\frac{p(1-p)}{N}}$, then the centre of the confidence interval is

$$\begin{aligned} &½\left(\frac{p_L}{1-p_L} + \frac{p_U}{1-p_U}\right) = ½\left(\frac{p-h}{1-p+h} + \frac{p+h}{1-p-h}\right) \\ &= \frac{p-\left(p^2-h^2\right)}{(1-p)^2-h^2} = \frac{p}{1-p} + \frac{h^2}{(1-p)\left((1-p)^2-h^2\right)}. \quad (2) \\ &\geq \frac{p}{1-p} = SR \end{aligned}$$

The centre is greater than the observed SR, but when $N \to \infty$, then $h \to 0$ and the centre converges towards SR.

The length of the CI is

$$CI_1 = \frac{p_U}{1-p_U} - \frac{p_L}{1-p_L} = \left(\frac{p+h}{1-p-h} + \frac{p-h}{1-p+h} \right)$$
$$= \frac{2h}{(1-p)^2 - h^2} = \frac{2k\sqrt{\frac{p(1-p)}{N}}}{(1-p)^2 - k^2 \frac{p(1-p)}{N}} \quad . \quad (3)$$
$$= \frac{2k}{(1-p)} \sqrt{\frac{p}{(1-p)N}} \left(\frac{1}{1-k^2 \frac{p}{(1-p)N}} \right)$$

Fellman and Eriksson [8] gave an alternative confidence interval for SR. According to the ML theory, the variance of SR is

$$Var(SR) = \left(\frac{df(p)}{dx} \right)^2 Var(p), \quad (4)$$

where $f(p) = SR = \frac{p}{1-p}$.

Hence,

$$Var(SR) = \left(\frac{d\left(\frac{p}{1-p} \right)}{dp} \right)^2 Var(p) \quad . \quad (5)$$
$$= \left(\frac{1}{(1-p)^2} \right)^2 \frac{p(1-p)}{N} = \frac{p}{(1-p)^3 N}$$

From (5), it follows that $SD(SR) = \frac{1}{1-p}\sqrt{\frac{p}{(1-p)N}}$ and the CI is

$$\left(SR - \frac{k}{1-p}\sqrt{\frac{p}{(1-p)N}}, SR + \frac{k}{1-p}\sqrt{\frac{p}{(1-p)N}} \right). \quad (6)$$

Obviously, the centre of the CI is SR. The length of the CI is

$$CI_2 = \frac{2k}{1-p}\sqrt{\frac{p}{(1-p)N}} \, . \quad (7)$$

In addition to the result that the centre of the Visaria CI converges towards SR, they obtained

$$\frac{CI_2}{CI_1} = \frac{2k\sqrt{\frac{p}{(1-p)N}} \left((1-p)^2 - k^2 \frac{p(1-p)}{N} \right)}{(1-p)2k\sqrt{\frac{p(1-p)}{N}}} \quad . \quad (8)$$
$$= \left(1 - k^2 \frac{p}{(1-p)N} \right) \leq 1$$

Hence, $CI_2 \leq CI_1$, but the ratio $\frac{CI_2}{CI_1} \to 1$ when $N \to \infty$. Hence, the CIs are asymptotically identical, and although the observed SRs are biased, both are applicable for large *N*.

As Visaria [9] pointed out, the CIs are crucial when differences in the SRs have to be interpreted. Fellman and Eriksson [8] presented Visaria´s and their CIs with respect to the sample size *N* given on a logarithmic scale. This figure is reprinted as Figure 1 in this study. We note that for small data sets the CIs are broad, and consequently, it is difficult to identify statistically significant differences. In addition, we observe that for small values of *N* there is a notable upward shift in Visaria's CIs. With increasing *N*, this shift vanishes. Fellman and Eriksson [8] presented also a new χ^2 test. Krackow et al. [10] presented a χ^2 test of the variations in the SRs based on the proportion of males.

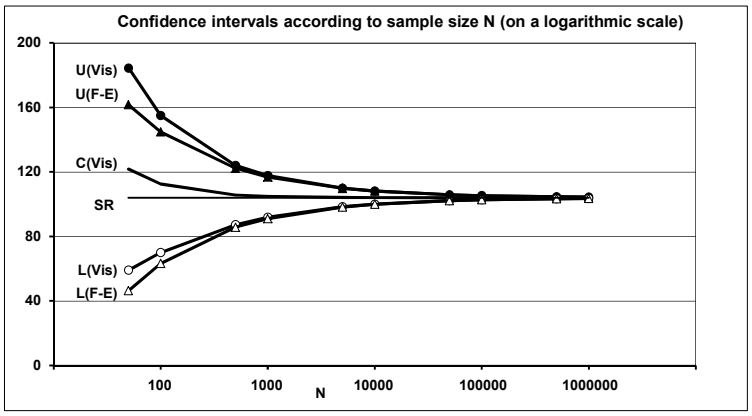


Figure 1. *Graphical sketch of the confidence intervals (CIs) for the secondary sex ratio (SR) according to the sample size N on a logarithmic scale. The CIs are given by Visaria [9] and by Fellman and Eriksson [8]. Upper (U) and lower (L) limits of the CIs are denoted U(Vis) and L(Vis) for Visaria and U(F-E) and L(F-E) for Fellman and Eriksson. The centre of the Visaria CI (C(Vis)) and the SR (104.0) are also included in the figure.*

Association between the stillbirth rate and the secondary sex ratio. Fellman and Eriksson [8] used the following notations: Let the number of males be $n(M)$, the number of females $n(F)$, the number of live-born males $n_L(M)$, the number of live-born females $n_L(F)$, the number of stillborn males $n_s(M)$ and the number of stillborn females $n_s(F)$. Consequently, the SRs are $SR_0 = \frac{n(M)}{n(F)}$ among all births,

$SR_L = \frac{n_L(M)}{n_L(F)}$ among the live-born and $SR_S = \frac{n_S(M)}{n_S(F)}$ among the stillborn. The stillbirth rate (SBR) among males is $SBR(M) = \frac{n_s(M)}{n(M)}$ and among females $SBR(F) = \frac{n_s(F)}{n(F)}$. The SR among the stillborn is

$$\begin{aligned} SR_s &= \frac{n_s(M)}{n_s(F)} = \frac{n_s(M)/n(F)}{n_s(F)/n(F)} \\ &= SR_0 \frac{n_s(M)/n(M)}{n_s(F)/n(F)} = SR_0 \frac{SBR(M)}{SBR(F)} \end{aligned}. \quad (9)$$

In general, $SR_0 \approx 1$, and hence, we obtain the general and simple, but approximate relation

$$SR_s \approx \frac{SBR(M)}{SBR(F)}. \quad (10)$$

Usually $SBR(M) > SBR(F)$ and $SR_S \geq SR_0 > 1$.

For the SR among the live-born, we obtain the formula

$$\begin{aligned} SR_L &= \frac{N_L(M)}{N_L(F)} = \frac{N(M) - N_s(M)}{N(F) - N_s(F)} \\ &= \frac{N(M)\left(1 - SBR(M)\right)}{N(F)\left(1 - SBR(F)\right)} = SR_0 \frac{\left(1 - SBR(M)\right)}{\left(1 - SBR(F)\right)} \end{aligned}. \quad (11)$$

In general, $SBR(M)$ and $SBR(F)$ are markedly less than one, and consequently, $\frac{1 - SBR(M)}{1 - SBR(F)} \approx 1$ and $SR_L \approx SR_0$. If $SBR(M) > SBR(F)$, then $SR_L < SR_0$.

Based on these analyses, $SR_S \geq SR_0 \geq SR_L$. Stillbirth rates are usually higher among males than females, and hence, the SR among stillborn infants is markedly higher than normal values, but the excess of males has decreased during the last decades. Hence, the SR among live-born infants is slightly lower than among all births, but this difference is today very minute.

Further, the SR among multiple maternities is lower than among singletons. The SR in multiple births is known to be low. The reason for this low ratio is not clear, but several hypotheses have been presented, including theories about maternal gonadotrophin levels at the time of conception or higher prenatal mortality of twin males [11 - 14].

In addition to these general findings, the SR shows marked regional and temporal variations. In a long series of papers, attempts have been made to identify factors influencing the SR, but statistical analyses have shown that comparisons demand large data sets.

Attempts to identify reliable associations between SRs and stillbirth rates have failed to yield consistent results. Hawley [15] stated that where prenatal losses are low, as in the high standard of living in Western countries, the SRs at birth are usually around 105 to 106. In the other hand, in low level-of-living areas where the frequencies of prenatal losses are relatively high, SRs at birth vary around 102. Visaria [9] could not find any correlation between the late foetal death ratios and the SRs of live births. Further, he stated that "the sex ratio at birth in the Negro population of the United States has not increased despite a marked fall in fetal mortality". Visaria concluded that there seems to be racial differences in the SR. Visaria [9] stressed that available data on late foetal mortality lend at best only weak support for these findings and concluded that racial differences seem to exist in the SR. He also discussed the perplexing finding that the SR among Koreans is high, around 113 [8].

3. Results

Temporal variations in the SR. In the 19th century, Berg [7] published SR data for live births in the counties of Sweden, 1749-1869 (c.f. Figure 4). The available periods varied between the counties. The regional data have been the basis for the study by Fellman and Eriksson [16]. They did not consider subperiod data for the counties because SR data for small data sets show large random fluctuations [8, 9]. Several studies have shown marked temporal variations in recent time series. A review of studies of temporal variations in the SR is given by Fellman and Eriksson [8]. For the period 1749-1869, such temporal trends seem to be minute.

Jalavisto [17] pointed out that in Finland the sex ratio has risen continuously and steadily from 1751 to 1948, and she showed that the lower rate of male births in the past could not be attributed to higher mean parity.

Recently, Møller [18] noted that the SR in Denmark showed an increasing trend up to the 1950s, followed by a decrease. He postulated that the initial increase in the male proportion is mainly a consequence of decreasing SBRs and the decreasing male excess among stillbirths. Møller's findings have inspired several authors investigating the temporal trends in the SR in different countries. They obtained similar results, but in different populations the locations of the peaks showed different patterns.

Bromen and Jöckel [19] studied German data for the period 1872-1995. They found clear peaks in the male proportion during the First and Second World War. Before and after the wars, the proportion was rather constant, but between the wars there was a marked trough. Van den Broek [20] compared the German data presented by Bromen and Jöckel [19] and Dutch data. He identified similar war effects as Bromen and Jöckel, but also found differences in the locations of the peaks. In a short comment, Parazzini et al. [21] stressed especially the decreasing trend in the proportion of males during the last 40 years.

Vartiainen et al. [22] evaluated the SR among live births in Finland, 1751-1997. They observed an increasing trend from 1751 to 1920; this was followed by a decrease and interrupted by peaks during and after World War I and II. None of the family parameters (paternal age, maternal age, age difference of parents and birth order) could explain the time trends. Furthermore, a brief rise occurred in the early 1970s. The last peak is more difficult to explain. The authors noted that the peak was associated with reduced birth rates and an increased

proportion of first-born children. Considering Jalavisto´s statement, these arguments do not convincingly explain the peak.

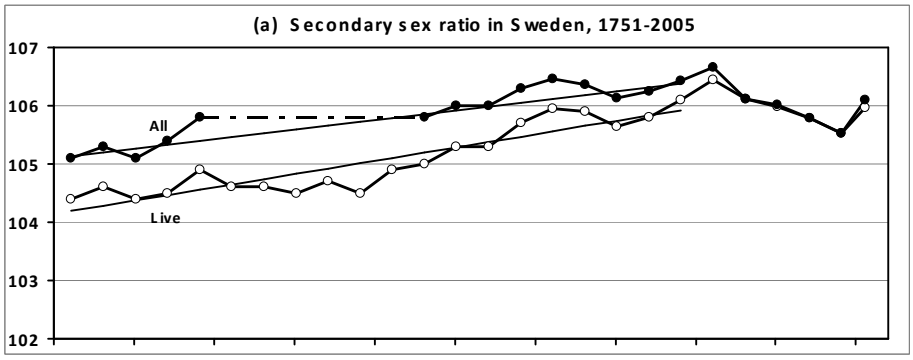
A common pattern observed in different countries is that during the first half of the 20th century the SR showed increasing trends, but during the second half the trend decreased. Different studies have found marked peaks in the proportion of males during the First and Second World War. It has been questioned whether temporal or spatial variations of the SR are evident, and whether they constitute an essential health event. A common opinion is that secular increases are caused by improved socio-economic conditions. The recent downward trends in the SRs have been attributed to new reproductive hazards, specifically exposure to environmental oestrogens. However, the turning point of the SR preceded the period of global industrialization and particularly the introduction of pesticides or hormonal drugs, rendering a causal association unlikely [23].

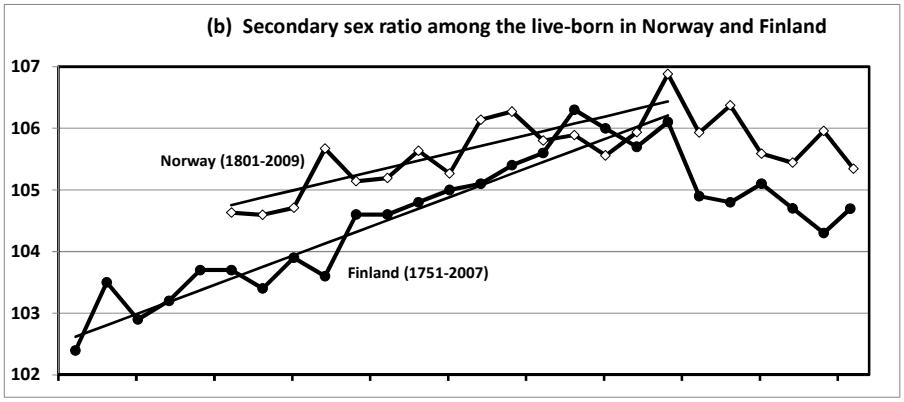
Fellman and Eriksson [23] compared the temporal trends in the SR in the Nordic countries and among all births (SR) and among live births (SR_L) in Sweden (1751-2005). The study is based on a combination of old data published by Berg [7, 24] and Sundbärg [25] and new data published by Statistics Sweden and "Socialstyrelsen". According to Eq. (11), the SBRs have only a very minor influence on SR_L . However, as noted above, some effects can be identified if long data series are analysed. During the period 1751-1960 the SR_L and the SR for all births in Sweden had unexplainable fluctuations, but they showed increasing tendencies. After 1960, both decreased. Fellman and Eriksson [16] obtained for SR_L a time series for the period 1751-2007, and for SR for the periods 1751-1800 and 1861-2004. Live births in Finland for the period 1751-2007 are given in the Statistical Yearbook of Finland (2008). They applied regression models for all births and live births in Sweden, 1751-1960, and for live births in Finland, 1761-1950. All models indicate statistically increasing trends. In Sweden, the more marked increasing trend in the SR for live births than for all births is obviously a result of the decreasing trend in the SBR.

Fellman and Eriksson [23] compared SR_L data for Finland and Norway. Increasing trends are observed for Finland for 1751-1950 and for Norway for 1801-1950. After the maxima, both series decreased. Included in the figures are the linear trends for the periods up to 1950. Their findings for Finland and Norway are in good agreement with the results for Sweden. Good agreement has also been found between the Danish and Icelandic series (*Icelandic Historical Statistics*. Reykjavík, 1997, 957 pp). In Finland and Denmark, the increasing trends are more pronounced than in Sweden and Norway. For Iceland, the trend is most marked. Figure 2 is a reprint of Figure 1 in Fellman and Eriksson [23].

Fellman and Eriksson [23] also presented the SBR for Sweden, Finland, Norway, Denmark and Iceland, and the temporal trends in all countries are very similar (Figure 3). They tried to identify associations between the SRs and the SBRs.

The association between the SRs and the SBRs is potentially disturbed by the fact that both may be influenced by external, time-dependent and yet unknown factors.





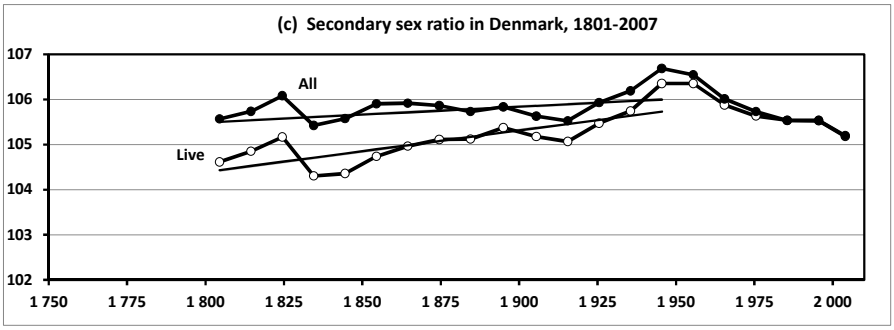


Figure 2. *Temporal trends in the secondary sex ratio (SR) (a) among live births and all births in Sweden, 1751-2005, (b) among live births in Finland (1751-2007) and Norway (1801-2009) and (c) among live births and all births in Denmark (1801-2007). Linear trends for the increasing SRs up to 1950 are included [23].*

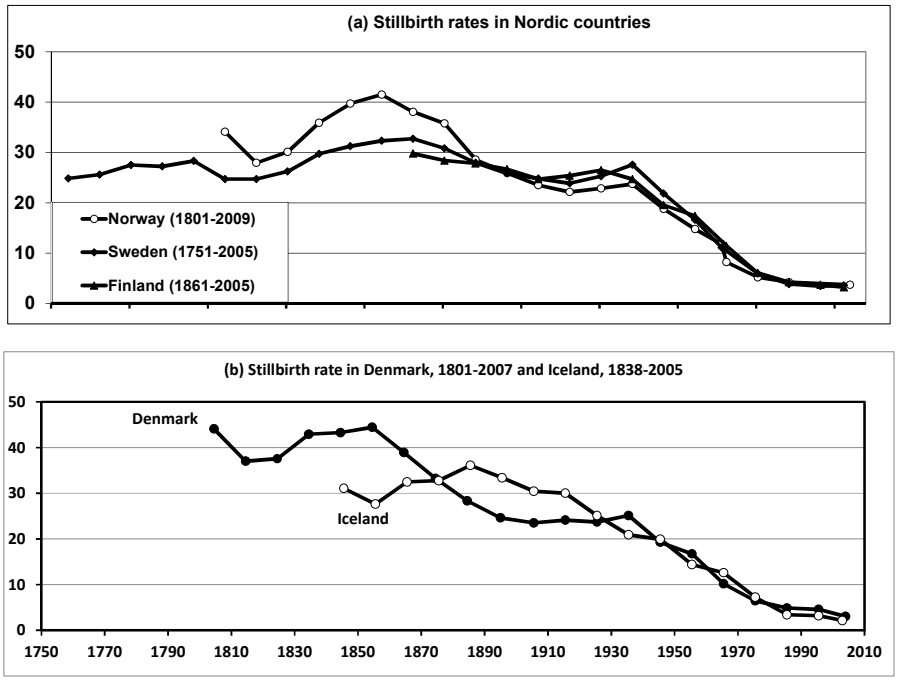


Figure 3. *Stillbirth rates per 1000 births in (a) Sweden (1751-2005), Norway (1801-2009) and Finland (1861-2007) and (b) Denmark (1801-2007) and Iceland (1838-2005). The decreasing trends are similar for all countries. The strong decreasing trends after 1940 and the disappearing sex differences in the SBRs cause a convergence between secondary sex ratio among all births and live births (cf. Sweden in Figure 1a and Denmark in Figure 1c) [23].*

Fellman and Eriksson [23] tried to remove the effect of these factors on the trends in the SR and SBR by considering time as a proxy variable for these factors. Consequently, they eliminated the effect of the unknown factors by studying partial correlation coefficients between SR and SBR when time is kept fixed.

The recent downward trends in the SRs have been interpreted in terms of new reproductive hazards, specifically exposure to environmental oestrogens. However, because the turning point of the SR precedes the introduction of pesticides or hormonal drugs, a causal association is rendered unlikely.

Furthermore, numerous authors have suggested that determinants of the male/female ratio may differ between countries. Visaria [9] found no socio-economic effect on the SR among the Afro-American population in the US. However, in our opinion, the secular increases up to 1950 in the SR observed in the Nordic populations are associated with improved socio-economic conditions. Fellman and Eriksson [23] agree with Møller [18] that this trend is a consequence of the SBR´s decrease, simultaneously with the decrease in the sex difference in the SBRs. However, the attempts to identify reliable associations between the SRs and the SBRs have yielded inconsistent results. The correlation between SR and SBR varies strongly when different periods are considered. The SRs show two trends, one increasing up to the 1950s and another decreasing after that. The SBRs are rather high and constant until 1950, thereafter decreasing. They calculated for these sub-periods the partial correlation coefficients between the SRs and SBRs, when time is eliminated. Following Kendall and Stuart [26], they used the formula $r_{SR,SBR.t} = \frac{r_{SR,SBR} - r_{SR,t} r_{SBR,t}}{\sqrt{\left(1 - r_{SR,t}^2\right)\left(1 - r_{SBR,t}^2\right)}}$ for the partial correlation between *SR* and *SBR* when time (*t*) is kept fixed. The approximate standard error is

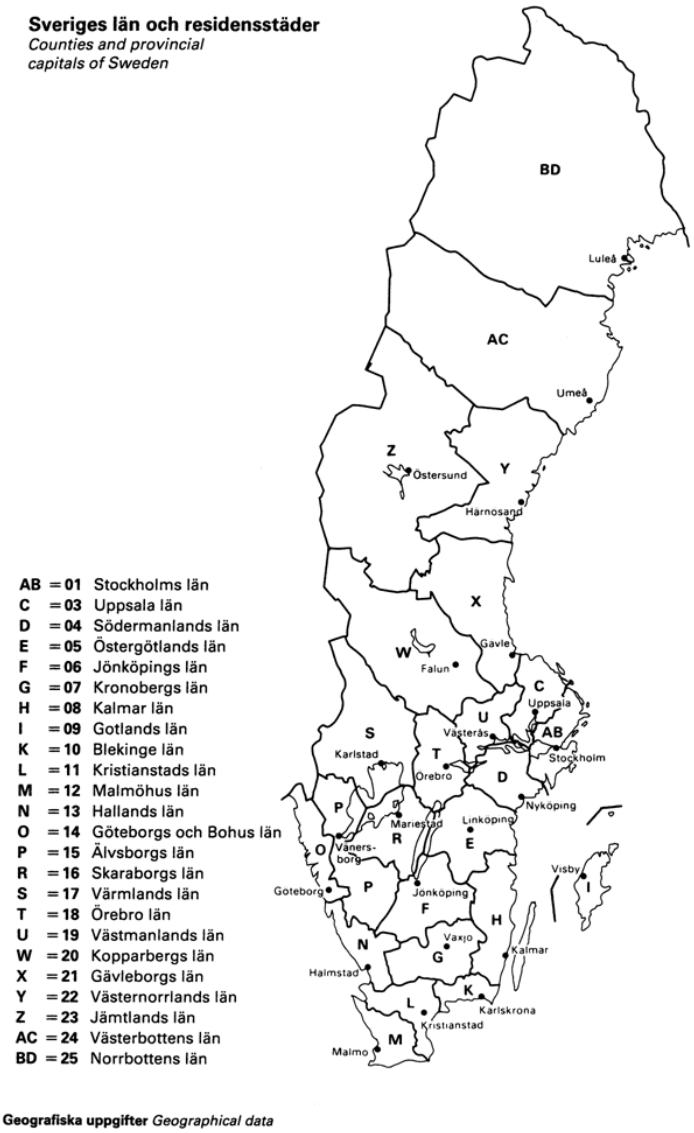
$$SE\left(r_{SR,SBR.t}\right) = \frac{1 - \rho_{SR,SBR.t}^2}{\sqrt{n}}.$$

From a theoretical standpoint, a better approximation would

be the use of the transformation $z = \frac{1}{2}\ln\left(\frac{1+r_{SR,SBR.t}}{1-r_{SR,SBR.t}}\right)$, but this transformed variable demands large data sets. According to the theory, the proposed influence of SBR on SR should produce negative correlation coefficients, but this is not generally the case. For periods up to 1950, the partial correlation coefficients were negative for Sweden (all and live births) and Denmark (live births). For the rest, the coefficients were positive, but close to zero. In recent decades, the correlation coefficients were, with the exception of Finland and Iceland, positive. Hence, there must be factors reducing the SR, despite the decreasing SBR. To sum up, universally consistent associations between SR and SBR were not found, and the negative results of Visaria [9] are confirmed.

Regional variations. Following Jalavisto [17] and Vartiainen et al. [22], Fellman and Eriksson [16] have connected the variations in the SR with two measures of fertility. They considered the crude birth rate (CBR) and the total fertility rate (TFR). Fellman and Eriksson [16] presented preliminary studies of the spatial variation of the variables

CBR, TFR, SR and TWR. Their results concerning SR are analysed and presented here.



8 **Geografiska uppgifter** *Geographical data*

Figure 4. *Map of Sweden including the counties (län) and their provincial capitals and the letter codes according to Statistics Sweden. The code AB includes both the city (A) and the county (B) of Stockholm [8].*

Hofsten and Lundström [27] stated that the boundaries of the counties in Sweden have only been subject to minor revisions, and consequently, the counties are ideal for use in analyses of geographical differences. The counties and their codes introduced by Statistics Sweden are presented in Figure 4 and used in Table 1. Berg [7] published SR data for live births in the counties of Sweden in 1749-1869 (Table 1). The available periods varied between the counties. Berg´s data have been transformed to number of males per 100 females and analysed and presented here. Hofsten and Lundström [27] presented in their Table 6.1 the CBRs for the counties in Sweden for the decades between 1751 and 1970. In this study, the variable CBR is the mean value of the decennial CBR data given by them for the period 1751-1870. Furthermore,

Hofsten and Lundström have in Tables 6.2 – 6.16 also presented TFR values per 1000 women for all decades starting from around 1860 to 1970. The variable TFR used in this study is their data for 1860.

In Table 1, we have included the regional data for SR, CBR, TFR and the number of live births (n) associated with the SRs.

Table 1 is an excerpt of Table 1 in Fellman and Eriksson [16] and displays the observation periods for the SR for the different counties. In accordance with the concepts outlined in their earlier study [28], Fellman and Eriksson introduced spatial regression models for the regional fertility data. The location of the counties was defined as the geographical co-ordinates of the corresponding residences (provincial capitals). The residences can be seen in Figure 4. They are not centrally located in the counties, but it is assumed that they are sufficiently central with respect to the population density, and their co-ordinates are given in Table 1. The presumptive regressors for the spatial regression models were the longitude (meridian) M and the latitude L and the transformed variables L^2, M^2 and LM. The regressors M and L were defined as deviations from the co-ordinates of the unweighted centre (59.18° N and 15.87° E) of the cluster of residences, and consequently, the intercepts obtained in the spatial models are the estimates of the regressands in this centre. The geographical co-ordinates for Sweden are eastern longitude and northern latitude. The elongated shape of Sweden indicates that attention must be paid to the multicollinearity between the regressors. Fellman and Eriksson [28] studied spatial models for the twinning rate in Sweden and provided a thorough analysis of the multicollinearity in the regression models with the geographical co-ordinates as regressors. In that study, they considered different measures of multicollinearity proposed in the literature. These were all functions of the eigen values (λ_i) of the correlation matrices. Fellman and Eriksson [16] restricted themselves to one measure, $m_4 = \sum_i \frac{1}{\lambda_i}$. This measure includes all eigen values, and for uncorrelated regressors, it equals the dimension of the correlation matrix, and with increasing multicollinearity, it increases towards infinity. For the total model containing all regressors L, M, L^2, M^2 and $L \cdot M$, they obtained the multicollinearity measure $m_4 = 38.09$. When all spatial regressors were included, the multicollinearity was rather strong, but for the optimal regression models with reduced number of regressors the multicollinearity measure was reduced.

Table 1. *Geographical co-ordinates, the number of live births associated with the SR, the secondary sex ratio (SR), the crude birth rate (CBR) and the total fertility rate (TFR) for the counties of Sweden. The residences are given in Figure 4. This table is an excerpt of Table 1 in Fellman and Eriksson [16].*

Codes[a)]	Periods[b)]	Lat.	Long.	n[c)]	SR	CBR[d)]	TFR[e)]
A	1749-1869	59.32	18.07	336854	103.4	34.5	3583
B	1749-1869	59.32	18.07	324901	104.6	31.4	4070
C	1749-1869	59.90	17.80	246343	104.2	30.7	4011
D	1749-1869	58.76	17.01	319940	104.6	31.0	4448
E	1749-1869	58.42	15.64	581692	104.5	32.4	4494
F	1749-1869	57.78	14.18	424184	104.9	31.6	4771
G	1749-1869	56.86	14.82	356405	104.7	34.2	4942
H	1749-1869	56.80	16.00	516021	105.8	33.7	4776
I	1749-1869	57.63	18.30	119541	105.3	28.2	3612
K	1749-1869	56.16	15.58	283511	103.7	35.6	4738
L	1749-1869	56.02	14.13	455200	104.8	32.8	4613
M	1749-1869	55.61	13.06	637249	104.6	34.9	4629
N	1749-1869	56.67	12.86	271859	104.5	32.0	4646
O	1749-1869	58.35	11.93	482251	103.9	34.8	4226
P	1749-1869	58.37	12.32	597113	104.8	32.5	4574
R	1749-1869	58.71	13.82	522657	104.8	33.5	5004
S	1749-1869	59.38	13.50	552016	105.1	33.0	4825
T	1749-1869	59.27	15.22	367499	104.5	32.3	5067
U	1749-1869	59.67	16.55	272943	104.3	31.4	4277
W	1749-1869	60.61	15.64	395484	104.7	30.5	4681
X	1810-1869	60.68	17.16	188398	104.8	29.0	4085
Y	1810-1869	62.63	17.94	175813	105.1	32.6	4880
Z	1810-1869	63.18	14.65	77473	106.4	27.3	4539
AC	1749-1869	63.83	20.27	169733	104.4	37.6	5366
BD	1749-1869	65.59	22.17	153877	105.1	37.6	5509
Total	1749-1869	59.18	15.87	8828958	104.66		

[a)] Codes are explained in Figure 4.
[b)] For Stockholm city and the county of Gotland, data are known for the whole period, but the rest of the counties have missing data for the period 1774-1794.
[c)] The number of live births for the defined period.
[d)] CBR is the mean value of the decennial CBR data given by Hofsten and Lundström [27].
[e)] TFR for 1860 given by Hofsten and Lundström [27].

Fellman and Eriksson analysed the spatial variation in the SR by weighted regression models. The regressand was the observed regional SR. The variance of the SR is approximately proportional to n^{-1}, and therefore, they used the number of live births (n in Table 1) in the counties as weights. The spatial variations in CBR and TFR were also studied with the geographic co-ordinates as regressors, but now no weights could be included in the regression analyses because no information about the heterogeneity in the variances was available. Fellman and Eriksson [16] tried to improve the spatial models for SR by including TFR and CBR as additional regressors.

The first analysis of the SR was to check the regional heterogeneity in the SR for live births. This was performed with χ^2 tests so that the numbers of males and females in the counties were estimated by the total number of live births and by published regional SRs, both given by Berg [7]. Significant regional differences in the sex proportions were found, and the next step was to build spatial models for the SR. In general, for moderate data sets, the SR is influenced by large random fluctuations [9, 23]. This can be seen in Figure 5, where the regional SRs are presented with 95% confidence intervals. Note the broad confidence intervals for the counties of Gotland (I), Gävleborg (X), Västernorrland (Y), Jämtland (Z), Västerbotten (AC) and Norrbotten (BD). For these, the number of live births is less than 200000.

The estimated optimal spatial model for SR derived in [16] is

$$SR = 104.56 + 0.0681\, L^2 - 0.0666\, L\, M \ .$$

The adjusted coefficient of determination $\overline{R}^2 = 0.103$ indicated a poor fit. The estimated parameter $\hat{\beta}_{L^2}$ is statistically significant, and $\hat{\beta}_{ML}$ is almost significant. The intercept $SR = 104.56$ is an estimate of the SR in a hypothetical county whose latitude equals 59.18° and longitude equals 15.87° . The most marked discrepancies between the observed and estimated SRs are seen in the city of Stockholm (A) and the counties of Blekinge (K), Kalmar (H) and Jämtland (Z) [29].

For the SR, Fellman and Eriksson [16] constructed a weighted regression model based on the spatial variables and the fertility variables CBR and TFR. In the optimal model, the spatial variables had insignificant parameter estimates and were eliminated. The fertility model was

$$SR = 104.68 + 0.000855\, TFR - 0.144483\, CBR \ .$$

The optimal model obtained has a rather good fit. The adjusted coefficient of determination was $\overline{R}^2 = 0.373$, and the regression parameter estimates were significant. Together with the observed SRs, the estimated SRs for the optimal model are given in Figure 6. The most marked discrepancies between the observed and estimated SRs are in the counties of Gotland (I), Kalmar (H) and Jämtland (Z).

No common spatial pattern for the demographic variables SR, TFR and CBR was detected, but a better fit was noted for TFR and CBR than for SR. Comparing these results, one observes that for the Eastern counties of Gotland (I), Uppsala (C) and Gävleborg (O) both fertility measures are low and for the Northern counties of Västerbotten and Norrbotten both measures are high. Hofsten and Lundström [27] reported that the CBR for the city of Stockholm (A) was above the CBR for the whole country, simultaneously with the TFR being low. They stressed that as early as about 1860 the city of Stockholm (low TFR in this study) and the county of Gotland (low TFR and CBR in this study) displayed a fertility considerably lower than that for the country overall. The difference being most marked in the higher age groups seems to indicate an early influence of birth control.

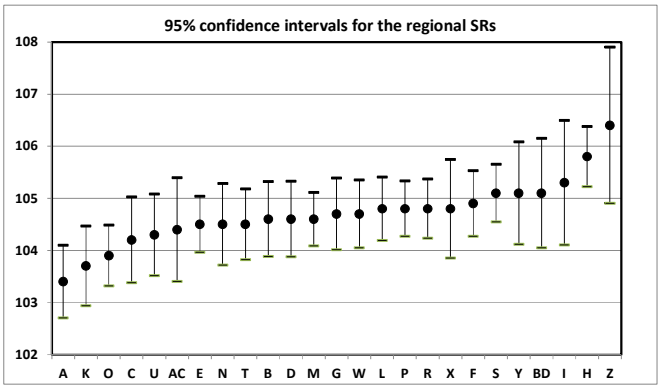


Figure 5. *Observed secondary sex ratios (SRs) for the period 1749-1869 and their confidence intervals (CIs) for different counties. Note the broad CIs for the counties of Gotland (I), Gävleborg (X), Västernorrland (Y), Jämtland (Z), Västerbotten (AC) and Norrbotten (BD). For these, the number of live births is less than 200 000. The counties are ordered according to increasing SR, and the county codes are given in Figure 4 [8].*

4. Discussion

Variations in the SR that have been reliably identified in family data have in general been slight and without notable influence on national birth registers (for references, see [30, 31]).

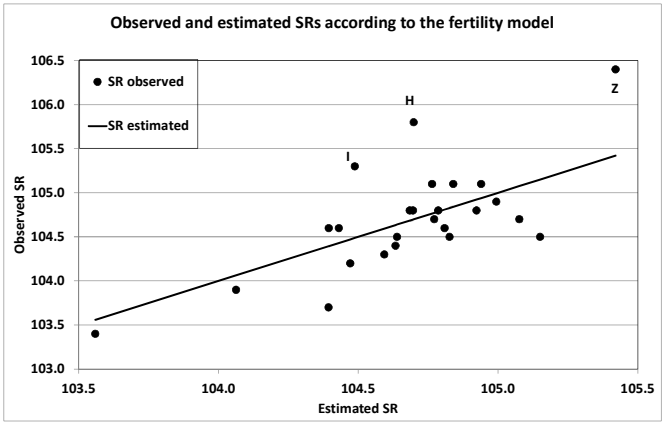


Figure 6. *Comparison between observed and estimated SRs in Sweden (1749-1869) according to the fertility model [23].*

In general, factors that affect the SR within families remain poorly understood. In a long series of papers, using family data, attempts have been made to identify factors influencing the SR. Increasing evidence confirms that exposure to chemicals, including pollutants from incinerators, dioxin, pesticides, alcohol, lead and other such workplace hazards, has produced children with a reduced male proportion. Variables reported to be associated with an increase in the SR are large family size, high ancestral longevity, paternal baldness, excessive coffee-drinking, intensive coital frequency and some male reproductive tract disorders.

Some striking examples can be found in the literature of unisexual pedigrees extending over several generations. Slater [32] stated that aberrant SRs tend, to some extent, to run in families. The finding by Lindsey and Altham [33] that the probability of couples being only capable of having children of one sex is very low contradicts Slater's statement.

The variation in the SR that has been reliably identified in family studies has consistently been slight compared with that observed in families with X-linked recessive retinoschisis (cleavage of retinal layers). A marked excess of males within such families was noted, in contrast to normal SRs in families with the X-linked recessive disorders haemophilia and colour blindness [30]. However, with the exception of the X-linked recessive retinoschisis, no unequivocal examples exist of genes in man that affect the SR, and X-linked retinoschisis is universally very rare. Summing up, while factors may have an effect on family data, they have not been identified in large national birth registers [30].

The familial variation in the SR that has been reliably identified has always been slight as compared with what has been noted in X-linked recessive retinoschisis [34].

In a long series of papers, attempts have been made to identify factors influencing the secondary sex ratio (SR). Hawley [15] stated that where prenatal losses are low, as in the high standard of living in Western countries, the SRs are usually around 105 to 106. On the other hand, in areas with a lower standard of living, where the frequencies of prenatal losses are relatively high, SRs are around 102. However, Visaria [9] stressed that available data on late fetal mortality lend at best only weak support for the hypothesis that the SR in birth registers varies inversely with the frequency of prenatal losses.

The downward trends have been interpreted in terms of new reproductive hazards, and specifically, to exposure to environmental oestrogens. Furthermore, Visaria suggested that determinants of the male/female ratio may differ between countries. However, the turning point of the SR precedes the period of industrialization or the introduction of pesticides or hormonal drugs, rendering a causal association unlikely.

In general, factors that affect the SR remain poorly understood. In a long series of papers, attempts have been made using family data to identify factors influencing the SR. Many of the causes of *reduced SR* that have been proposed or identified, such as stress of fathers, in vitro fertilization, less frequent intercourse and multiple sclerosis, are unlikely to account for the temporal trends in the SRs that have been observed in several industrial countries. Increasing evidence confirms that exposure to chemicals, such as pollutants from incinerators, dioxin, pesticides, alcohol, lead and other such workplace hazards, have produced children with a reduced male proportion. Moreover, while some male reproductive tract disorders have increased, leading to the questions of whether temporal or spatial variations in the SR are evident and if so whether they constitute an essential health event [35].

Increased SRs have been reported for some human malformations [36] and in some parental disorders [37 - 39]. Other variables reported to be associated with an increase in the SR are large family size, high ancestral longevity, paternal baldness, excessive coffee-drinking, intensive coital frequency and illegitimacy [40]. However, Fellman and Eriksson [8] could not confirm any effect of marital status on the SRs observed in national birth registers.

Biased SRs have also been noted in connection with neural tube defects [41]. James [38, 42] hypothesized that the hormone levels of both parents at the time of conception affect the probability of a male birth, high levels of oestrogen and testosterone increasing this probability, and high levels of gonadotrophin decreasing it.

Genetic control of the SR could occur either through control of the primary sex ratio or through subsequent differential sex mortality. The only clear-cut case of distortion of segregation, or meiotic drive, in man involves the D-21 translocation, which is implicated in a certain proportion of cases of Down syndrome. Normal carrier males rarely transmit the D-21 combination to their offspring (only about 5% of that expected). This example clearly shows that genetic variation can affect the segregation ratio.

Examples can be found in the literature of unisexual pedigrees extending over several generations, which, if valid, suggest the existence of SR genes like those reported in Drosophila [43, 30]. Slater [32] stated that aberrant SRs tend, to a slight extent, to run in families, but to our knowledge no further studies have been carried out to confirm this observation.

There may be a disturbance in the primary sex ratio, possibly due to meiotic drive (primary segregation disturbance or genetic selection), resulting in overproduction or preference of the egg for Y sperm. Some indications for this come from the observation of an unusually high proportion of males in Koreans (proportion of males somewhat above 0.53 and SR ratio about 113). However, Korean mothers have a normal SR among their progeny, ruling out effects of sex-differential mortality in utero [44].

However, with the exception of the X-linked recessive retinoschisis confirmed with data expanded and more advanced statistical methods [30, 45], there are no unequivocal examples of genes in man that affect the SR. To the best of our knowledge, the only genetic characteristic observed to have a marked influence on the SR within families is X-linked retinoschisis, but this disorder is universally very rare. Summing up, all of these above-mentioned factors, although they may have an influence on family data, are not identifiable in large national birth registers.

Visaria concluded that racial differences appear to exist in the SR. He also noted that the observed SRs are strongly influenced by random errors, and he constructed confidence intervals (CIs) for the SR. Fellman and Eriksson ([8]) presented a new formula for the CI and analysed in detail the SR among singleton, twin and higher multiple births in Sweden (1869-1967). These CIs are asymptotically the same, and are applicable for large *N*.

Fellman and Eriksson [23] compared the temporal trends in the SR among all births (SR_A) and among live births (SR_L) for Sweden. We obtain for SR_L a time series for the period 1751-2008, and for SR_A for the periods 1751-1800 and 1861-2007. The study is based on a combination of old birth data published by Sundbärg [25] and new data published by Statistics Sweden. During 1751-1960, the SR_L and SR_A had unexplainable fluctuations, but showed increasing trends. To emphasize the differences in the trends, we include trend lines for both series for 1751-1960. The more marked increasing trend in the SR_L than in the SR_A is obviously a result of the decrease in the SBR. After the 1960s, the SR_A and SR_L converge and show similar decreasing trends. This pattern is caused by the decrease in the SBR.

Trivers and Willard [46] report that theory and data suggest that a male in good condition at the end of the period of parental investment is expected to outreproduce a sister in similar condition, while she is expected to outreproduce him if both are in poor condition. Accordingly, natural selection should favour parental ability to adjust the sex ratio of offspring produced according to parental ability to invest. Data from mammals support the model; as maternal condition declines, the adult female tends to produce a lower ratio of males to females.

Sieff [47] presented an extensive review, including references, concerning factors influencing the SR. Increasing evidence confirms that exposures to chemicals, such as pollutants from incinerators, dioxin, pesticides, alcohol, lead and other such workplace hazardous substances produced children with a reduced male proportion, while some male reproductive tract disorders have increased. It has been questioned whether temporal or spatial variations of the SR are evident, and whether they constitute an essential health event [35].

Recently, Garenne [48] evaluated classical family data [49, 50] and the straightforward data analysis revealed a strong family heterogeneity in the SR.

Acknowledgement

This study was in part supported by a grant from the foundation “Magnus Ehrnrooths Stiftelse”.

References

[1] Graunt, J. (1662) Natural and Political Observations on the Bills of Mortality.

[2] David, F. N. (1962). Games, Gods and Gambling. Charles Griffin & Co. Ltd, London. xvi+275 pp.

[3] Arbuthnot, J. (1711). An Argument for Divine Providence, taken from the Constant Regularity observed in the Births of both Sexes.

[4] Campbell, R. B. (2001). John Graunt, John Arbuthnott, and the human sex ratio. Human Biology, 73:605-610.

[5] Hacking, (1975). The Emergence of Probability. Cambridge University Press, Cambridge 209 pp.

[6] Fellman, J. (2011). Sex ratio at birth. In International Encyclopedia of Statistical Science, (Ed. Miodrag Lovric) Part 19:1314-1316

[7] Berg F. T. (1871). Proportionen mellan könen bland de födde och inom den stående befolkningen med hänsyn till Sverige och dess provinciela olikheter. (Sex ratio at birth and in the population with respect to Sweden and its regions). Kungliga Svenska Vetenskaps-Akademiens Handlingar 10(6):1-40.

[8] Fellman, J. & Eriksson, A. W. (2010). Secondary sex ratio in multiple births. Twin Research and Human Genetics, 13(1):101-108.

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

[10] Krackow, S., Meelis, E. & Hardy, I. C. W. (2002). Analysis of sex ratio variances and sequences of sex allocation. In I. C. W. Hardy (Ed.). Sex Ratios Concepts and Research Methods (pp. 112–131). Cambridge, UK: Cambridge University Press.

[11] Milham, S. (1964). Pituitary gonadotrophin and dizygotic twinning. Lancet, 2, 566.

[12] Bulmer, M. G. (1970). The Biology of Twinning in Man. London: Oxford University Press.

[13] James, W. H. (1980). Time of fertilization and sex of infants. Lancet, 24,:1124–26.

[14] James, W. H. (1986). Hormonal control of sex ratio. J. Theor. Biol. 118:427-431.

[15] [Hawley, A. H. (1959). Population composition. In: The Study of Population: An Inventory and Appraisal. Ed. P. M. Hauser and O. Dudley Duncan. Chicago, University of Chicago Press, 361-382.

[16] Fellman, J. & Eriksson, A. W. (2014). Effect of fertility on secondary sex ratio and twinning rate in Sweden, 1749-1870. Twin Research and Human Genetics 18, 2014:92-99.

[17] Jalavisto, E. (1952). Sex ratio at birth and its dependence upon birth order and parental age. Ann. Chir. Gynaec. Fenn., 41:129-137.

[18] Møller, H. (1996). Change in male:female ratio among newborn infants in Denmark. Lancet, 348:828-829.

[19] Bromen, K. & Jöckel, K.-H. (1997). Change in male proportion among newborn infants. Lancet, 349:804-805.

[20] van den Broek, J. M. (1997). Change in male proportion among newborn infants. Lancet, 349:805.

[21] Parazzini, F., La Vecchia, C., Chatenoud, L., Chiaffarino, F. & Benzi, G. (1997). Change in male proportion among newborn infants. Lancet, 349:805-806.

[22] Vartiainen, T., Kartovaara, L. & Tuomisto, J. (1999). Environmental chemicals and changes in sex ratios: Analysis over 250 years in Finland. Environmental Health Perspectives. 107:813-815.

[23] Fellman, J. & Eriksson, A. W. (2011). Temporal trends in the secondary sex ratio in Nordic countries. Biodemography and Social Biology 57:2: 143-154.

[24] Berg, F. T. {1880). Om flerfostriga barnsbörder (On multiple maternities, in Swedish) Hygiea (Stockholm) 42, 331-342.

[25] Sundbärg, G. (1907). Bevölkerungsstatistik Schwedens 1750-1900. Einige Hauptresultate (Swedish population statistics. Main results). P. A. Norstedt & Söner, Stockholm. 1907, 170pp.

[26] Kendall, M. G. & Stuart, A. (1967). The Advanced Theory of Statistics: Vol. 2. Chapter 26. (2nd ed., pp 690). London: Charles Griffin.

[27] Hofsten, E. & Lundström, H. (1976). Swedish Population History. Main trends from 1750 to 1970. Urval 8. National Central Bureau of Statistics. Stockholm, 186 pp.

[28] Fellman, J. & Eriksson, A. W. (2009). Spatial variation in the twinning rate in Sweden, 1751-1850. Twin Res. Hum. Genet. 12(6):583-590

[29] Fellman, J. & Eriksson, A. W. (2015). Geographic variation in fertility measures in Sweden in (1749-1870). British Journal of Medicine & Medical Research, 7(1):1-10.

[30] Fellman J., Eriksson A. W. & Forsius H. (2002). Sex ratio and proportion of affected sons in sibships with X-chromosomal recessive traits: Maximum likelihood estimation in truncated multinomial distributions. Human Heredity, 53:173-180.

[31] Fellman, J. & Eriksson, A. W. (2008). Sex ratio in sibships with twins. Twin Res. Hum. Genet. 11:204-214.

[32] Slater E. (1943). A demographic study of a psychopathic population. Annals of Eugenics, 12:121-137.

[33] Lindsey, J. K. & Altham, P. M. E. (1998). Analysis of the human sex ratio by using overdispersion models. Applied statistics, 47:149-157.

[34] Eriksson, A.W., Vainio-Mattila, B., Krause, U., Fellman, J. & Forsius, H. (1967). secondary sex ratio in families with X-chromosomal disorders. Hereditas, 57:373-381.

[35] Davis, D. L., Gottlieb, M. B. & Stampnitzky, J. R. (1998). Reduced ratio of male to female births in several industrial countries. JAMA, 279:1018-1023.

[36] Arena, J. F. P. and Smith, D. W. (1978). Sex liability to single structural defects. Amer. J. Dis. Child. 132:970-972.

[37] James, W. H. (1987a). The human sex ratio. Part 1: A review of the literature. Hum. Biol. 59:721-752.

[38] James, W. H. (1987b). The human sex ratio. Part 2: A hypothesis and a program of research. Hum. Biol. 59:873-900.

[39] Møller, H. (1998). Trends in sex-ratio, testicular cancer and male reproductive hazards: Are they connected? APMIS: Acta Path. Microbiol. Immunol. Scand. 106:232-239.

[40] Teitelbaum, M. S. (1972). Factors associated with the sex ratio in human populations. In: The Structure of Human Populations pp. 90-109. Edited by Harrison GA and Boye AJ. Clarendon Press, Oxford.

[41] Seller, M. J. (1987). Neural tube defects and sex ratios. Amer. J. Med. Genet. 26:699-707.

[42] James, W. H. (1996). Evidence that mammalian sex ratios at birth are partially controlled by parental hormone levels at the time of conception. J. Theor. Biol. 180:271-286.

[43] Stern, C. (1960). Principles of Human Genetics, 2nd ed. W H Freeman, San Francisco.

[44] Morton, N., Chung, C. S. & Mi, M.-P. (1967). Genetics of Interracial Crosses in Hawaii. Monographs in Human Genetics Vol. 3. Ed´s Beckman L and Hauge M. S. Karger Basel. 160 pp.

[45] Huopaniemi, L., Fellman, J., Rantala, A., Eriksson, A., Forsius, H., de la Chapelle, A. & Alitalo, T. (1999). Skewed secondary sex ratio in the offspring of carriers of the 214G>A mutation of the RS1 gene. Annals of Human Genetics 63:1999: 521-533. In Laura Huopaniemi: Molecular Genetics of X-Chromosomal Juvenile Retinoschisis, Helsinki 2000 (diss)

[46] Trivers, R. L. & Willard, D. E. (1973). Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring. Science, 179, 4068:90-92.

[47] Sieff, D. F. (1990). Explaining biased sex ratios in human populations. Current Anthropology, 31:25-48.

[48] Garenne, M. (2009). Heterogeneity in the sex ratio at birth in European populations. Genus, 64:99-108.

[49] Geissler, A. (1889). Beiträge zur Frage des Geschlechtverhältnisses der Geborenen (Contributions to the secondary sex ratio). *Z. Königl. Sächs. Statist. Bur. 35*:1-24.

[50] Malinvaud, E. (1955). Relations entre la composition des familles et le taux de masculinité. *J. Soc. Statist. Paris*, *96*:49.