

Spectrum of Sex Ratios in Denmark

Abstract. The sex ratio (SR) is usually defined as the number of males per 100 females within an area or, as in this study, the proportion of males among all births (P_M). It has been observed that among newborns, there is typically a slight excess number for boys compared to girls. Consequently, the SR becomes greater than 100, which is around 106 in number, and the chance of new born males is around 0.515. Attempts have been made to identify the factors those are influencing the level of the P_M . Previous researches stated that where prenatal losses are low, as in the Western countries, the SRs are also become high around 105 to 106, but in areas where the frequencies of prenatal losses are relatively high then the SRs are found to be low around 102. Later on several researches have focused on temporal, regional and seasonal fluctuations of SR. In general, factors that affect the SR within the families remain poorly understood. Attempts to identify such factors in national birth registers are also remained to be unsuccessful. Recently, SR studies have mainly concentrated on the identification of general but occasional factors. In this study, we tried to identify the effects of issues like maternal age and type of delivery (live- and stillborn, singletons and multiples) to identify the controlling parameters of sex ratio during birth. Post experimental outcome showed that there is no significant difference between live- and stillborn and maternal age had as no significant effect for controlling sex ratio. The SR is higher among singletons than that of multiples, but there is no significant difference obtained in SR between twins and triplets. Among singletons the temporal differences are non-significant, but for twins and triplets, significant temporal differences were obtained.

Keywords: Sex ratio, Maximum likelihood; confidence intervals; temporal trends; maternal age; multiple maternities

Introduction

The sex ratio (SR) at birth is usually defined as the number of males per 100 females, or the ratio of males to females, or, as in this study, the proportion of males among all births (P_M). Among newborns, there is typically a slight excess of boys present at every population. Consequently, the SR is greater than 100, mainly around 106 and the P_M is around 0.515. In the 17th century, lots of discussion had been carried out to identify the reasons behind this excessiveness. One argument was that this balance between male and female births is not the

33 result of chance but Divine Providence. Opponents argued that even if the chance of a male,
 34 P_M , differs from 0.5 sufficiently, repeated trials would result in a relative frequency very
 35 close to P_M . Many descriptions are elaborated at various studies regarding this uneven rates
 36 of males and females [1] [2]. Today, the study of secondary sex ratio is based on a more
 37 stable foundation of statistical theory, which enables scientists. to identify the influential
 38 factors. The interest has mainly focused on temporal, regional and seasonal fluctuations of
 39 SR. Temporal fluctuations are often been connected to war and peace. No family parameters
 40 are able to explain the time trends. In general, factors that affect the SR within families
 41 remain poorly understood. Attempts to identify such factors from national birth registers have
 42 failed [3]. Recently, SR studies have concentrated on the identification of general but
 43 occasional factors. Grech and Borg tried to identify such factors in a series of their studies
 44 [4]. In this paper, we studied the sex ratio using the variable P_M . The data for this study were
 45 collected over periods, too short to allow the effects of temporal factors to be determined.

46 **Materials and Methods**

47 **Maximum likelihood estimation.** If the theoretical proportion of males is p_0 , then the
 48 observed relative frequency of males p is a maximum likelihood (ML) estimator of p_0 , being
 49 unbiased, consistent, efficient and asymptotically normal with $E(p) = p_0$ and
 50 $\text{Var}(p) = p_0(1 - p_0)/N$. This finding speaks in favour of the use of P_M as a parameter for sex
 51 ratio measurement. According to the ML theory, the sex ratio $\text{SR} = p/(1 - p)$ is a ML
 52 estimator of the transformed parameter $\text{SR}_0 = p_0/(1 - p_0)$, but SR is not unbiased. When
 53 $N \rightarrow \infty$, then $p \rightarrow p_0$ and $\text{SR} \rightarrow \text{SR}_0$, and the estimated SR is consistent, biased, but
 54 asymptotically unbiased, and normally distributed [5].

55 **Standard deviations and confidence intervals.** Visaria [6] pointed out that random errors
 56 influence the variation in the SR. Therefore, he presented a numerical table about how the
 57 confidence intervals (CIs) of the SR depend on the observed SR and the number of births. He
 58 gave no formula for the intervals, but stated that “the standard error of an observed sex ratio
 59 can be estimated as the standard error of the proportion p of male births among the total”[6].
 60 Fellman and Eriksson [5] interpreted Visaria’s statement such that he constructed CIs for p ,
 61 that is

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

where k corresponds to the confidence level. After that, Visaria defined the CI for the SR,

(SR_L, SR_U) , so that $SR_L = p_L/(1-p_L)$ and $SR_U = p_U/(1-p_U)$, where

$$p_L = p - k\sqrt{\frac{p(1-p)}{N}} \text{ and } p_U = p + k\sqrt{\frac{p(1-p)}{N}}. \quad (2)$$

Visaria's attempt is based on the fact that SR is a monotonously increasing function of p .

The length of Visaria's CI is as follows:

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

Fellman and Eriksson [5] gave an alternative confidence interval for SR. According to the ML theory, the variance of SR is as follows:

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

where $f(p) = \text{SR} = p/(1-p)$. Hence,

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

From (5), it follows that the standard deviation (SD) of SR is

$$\text{SD}(\text{SR}) = \frac{1}{(1-p)} \sqrt{\frac{p}{(1-p)N}}$$

and the CI is

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

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

$$\text{CI}_2 = \frac{2k}{(1-p)} \sqrt{\frac{p}{(1-p)N}} \quad (7)$$

In addition to the finding that the centre of the CI converges towards SR as observed by Visaria, Fellman and Eriksson [5] obtained

$$\frac{\text{CI}_2}{\text{CI}_1} = 1 - k^2 \frac{p}{(1-p)N} \leq 1. \quad (8)$$

Hence, $\text{CI}_2 \leq \text{CI}_1$, but the ratio $\text{CI}_2/\text{CI}_1 \rightarrow 1$ when $N \rightarrow \infty$. Thus, the CIs are asymptotically identical, and although the observed SRs are biased, both are applicable for large N .

Statistical tests: Fellman and Eriksson [5] presented also a new χ^2 test. The simultaneous tests of several SRs results in χ^2 test. Let the null hypothesis be that the sets have a common SR_0 . According to Fellman and Eriksson (5), the variance is

$$\text{Var}(\text{SR}_t) = p_0 / \left[(1-p_0)^3 N_t \right],$$

where $p_0 = \text{SR}_0 / (1 + \text{SR}_0)$ and N_t is the sample size for $t = 1, \dots, T$ and $T \geq 2$. The variable $(\text{SR} - \text{SR}_0) / \sqrt{p_0 / \left[(1-p_0)^3 N_t \right]}$ is a standardized variable, and consequently, asymptotically as $N(0, 1)$. Hence,

$$\chi^2 = \frac{(1-p_0)^3}{p_0} \sum_{t=1}^T N_t (\text{SR}_t - \text{SR}_0)^2, \quad (9)$$

χ^2 is asymptotically distributed with T degrees of freedom (DF). If p_0 is unknown, we introduce the weighted mean $\overline{\text{SR}} = \sum_{t=1}^T (N_t \text{SR}_t) / N_t$, which under the null hypothesis is the most efficient estimate of SR_0 , and estimate $p_0 = \overline{\text{SR}} / (1 + \overline{\text{SR}})$. Consequently, we have estimated one parameter, and

$$\chi^2 = \frac{1}{\bar{p}_0} (1 - \bar{p}_0) \sum_{t=1}^T N_t (\text{SR}_t - \overline{\text{SR}})^2 \quad (10)$$

χ^2 is distributed with T-1 degrees of freedom.

Krackow et al. [7] presented an analogous χ^2 test based on the proportion of males. Their method followed the same ideas as the Fellman and Eriksson's test based on the SR. Although we observed that sufficient statistical methods exist for the analyses of the SR, it may be noted that the proportion of males' birth among the total (P_M) is a more convenient alternative for the statistical study of the proportions of males and females. We therefore consider this variable herewith.

Materials. This study is based on data sets that have been obtained from the official Danish State Bank produced by Statistics Denmark. Some of these data were submitted for this study by Caroline Østerholm-Jørgensen of Statistics Denmark. The goal of our investigations is to examine the effects of different possible factors which are responsible in generating the views of number differences of males and females in a population. Consequently, we have to base our studies on different Danish data sets. Denmark is a small country and in order to obtain sufficiently large data sets, especially data concerning multiple maternities, the data must be collected over long periods. Accordingly, one must investigate whether there are temporal effects on the SRs.

Results

Live- and stillborn. In Fig. 1, we compare the temporal trends of P_M for live- and stillborn children. When we test temporal effects of P_M on live births, we obtain $\chi^2 = 5.16$ with 6 DF, and the temporal effect is non-significant ($P > 0.05$). For the stillborn, the corresponding test results are $\chi^2 = 8.38$ with 6 DF and $P > 0.05$. When we compare the total data sets for live-born and stillborn children we observe that the difference between the total means is $z = 0.00433$ with the $z = 0.00433$ and $t = 0.581$, and consequently, the difference is non-significant ($P > 0.05$). For the total data set, $P_M = 0.51305$ with $SD = 0.000445$.

Figure 1

Maternal age. We studied the effect of maternal age based on Danish data for the period of 1973-2016. The results are presented in Figure 2. No significant maternal age effect was

identified: ($\chi^2 = 3.299$ with 6 DF and $P > 0.05$). The total mean is $P_M = 0.5133$ with SD = 0.000302. Significant temporal differences were not obtained. In Fig. 2, a 95% confidence band is included. It can be noted how the size of the data set influences the band.

Figure 2

Singletons and multiples. Based on Danish data for the period of 1997–2016, we can compare the temporal trends of P_M among singletons and multiples. For singletons, we obtain $\chi^2 = 3.437$ with 6 DF, and the temporal differences were insignificant ($P > 0.05$) of its own. For multiple maternities, $\chi^2 = 9.969$ with 6 DF and $P > 0.05$. When we compare singletons and multiples for the total data sets, we observe that the mean difference is 0.0055 with SD = 0.00158 and $t = 3.49$, and consequently, the difference is found to be significant ($P < 0.001$), and the P_M is higher among singletons than multiples (Fig. 3).

Figure 3

Twins and triplets. Based on Danish data for the period of 1911 – 2016, we can compare the temporal trends of P_M among twins and triplets (Fig. 4). For twins, we obtain $\chi^2 = 43.71$ with 20 DF, and the temporal differences are found to be significant ($P < 0.01$). For triplets, the test result shows $\chi^2 = 56.134$ with 20 DF and $P < 0.001$. When we compare twins and triplets for the total data sets, we observe that the mean difference is 0.0007057 with SD = 0.007923 and the difference is non-significant. Hence, there is no significant difference between P_M among twins and triplets. For the total set of data it is found that, $P_M = 0.509568$ with SD = 0.001066. Furthermore, we observe that there is no significant correlation between time and P_M for either twins ($r = -0.0410$) or triplets ($r = 0.1168$).

Figure 4

Discussion

From the 17th century, many discussions ensued on the causes of the excess of boys over girls but no potential outcomes are obtained in this regard.. Berg [8] presented a detailed analysis of the SR in Sweden. 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”. Attempts have later been made to identify factors influencing the level of the SR. Hawley [9] stated that where prenatal losses are low, as in Western countries, the SRs are usually around

105 to 106, but in areas where the frequencies of prenatal losses are relatively high the SRs are around 102. Visaria [6] pointed out that the CIs are crucial when differences in the SRs are interpreted. Fellman and Eriksson [5] presented Visaria's depiction and their CIs with respect to the sample size N given on a logarithmic scale. They also noted that for small data sets the CIs are broad, and consequently, it is difficult to identify statistically significant differences.

Statistical analyses have shown that comparisons between SRs demand large data sets because random fluctuations in moderate data are marked. Consequently, reliable results presuppose the national birth data. The interest has focused on temporal, spatial and seasonal fluctuations of the SR. Temporal trends in SR were obtained especially in the Nordic countries, including Denmark, and linear trends for increasing SRs up to 1950 have been noted [10], [11].

Fellman and Eriksson [12] analysed the effect of total fertility rate (TFR) and crude birth rate (CBR) on the SR. They analysed historical demographic data and the regional variations of the counties in Sweden, and they built spatial models and as regressors they used geographical coordinates for the provincial capitals of the counties. The SR among the live born in 1749–1869 and the twinning rate (TWR) in 1751–1860 showed slight spatial variations. The influence of CBR and TFR on the SR and TWR was examined, and statistically significant effects were found.

Conclusion

It can be stated that the reasons behind the variation of Sex Ratios still remain unsolved. Fellman [13] noted a slight increasing trend of SR in Sweden from 1751-1870. This is probably associated with the country's increasing standard of living. A minor trough was observed between 1800 and 1860. This result combined with other evidence points to a mild subsistence crisis in western Sweden during this period. In general, factors that affect the SR within families of such population remain poorly understood. Attempts to identify such factors in national birth registers have been unsuccessful.

References

- [1] Fellman, J. Glimpses at the History of Sex Ratio Studies. *Sci J of Public Health* 5a (2015a) 291-302.
- [2] Fellman, J. Some aspects of sex ratio studies, *Acta Paediatrica* 105 (2016) 1371.

- [3] Fellman, J. Sex ratio at birth. In *International Encyclopedia of Statistical Science*, (Ed. Miodrag Lovric) *Part 19* (2011) 1314-1316
- [4] Grech V. and Borg T. Seasonality of sex ratio at births in Iceland and effects of the 2010 Eyjafjallajökull volcanic eruption. *Acta Paediatrica* 105 (2016) 369–70.
- [5] Fellman, J. & Eriksson, A. W. Secondary sex ratio in multiple births. *Twin Res. and Hum. Genetics*, 13(1) (2010) 101-108.
- [6] Visaria P. M. Sex ratio at birth in territories with a relatively complete registration. *Eugenics Quarterly* 14 (1967) 132-142.
- [7] Krackow, S., Meelis, E. & Hardy, I. C. W. Analysis of sex ratio variances and sequences of sex allocation. In I. C. W. Hardy (Ed.). *Sex Ratios Concepts and Research Methods* Cambridge, UK: Cambridge University Press 2002 pp. 112–131.
- [8] Berg, F. T. Proportionen mellan könen bland de födde och inom den stående befolkningen med hänsyn till Sverige och dess provinciella olikheter (Sex ratio at birth and in the population with respect to Sweden and its regions). *Kungl Svenska Vetenskapsakademiens Handlingar*, 10(6) (1871) 1–40.
- [9] Hawley, A. H. Population composition. In: *The Study of Population: An Inventory and Appraisal*. Ed. P. M. Hauser and O. Dudley Duncan. Chicago, University of Chicago Press, 1959 pp. 361-382.
- [10] Fellman, J. & Eriksson, A. W. Temporal trends in the secondary sex ratio in Nordic countries. *Biodemography and Soc. Biology* 57:2 (2011) 143-154.
- [11] Fellman, J. Temporal and regional trends in the secondary sex ratio: the Nordic experience. *Early Hum. Development*, 91 (2015c) 801–806.
- [12] Fellman, J. & Eriksson, A. W. Effect of fertility on secondary sex ratio and twinning rate in Sweden, 1749-1870. *Twin Res. and Hum. Genetics*, 18 (2014) 92-99. doi:10.1017/thg.2014.71.
- [13] Fellman J. Regional variation in the secondary sex ratio in Sweden, 1749-1869. *Br. J. of Appl. Sci. & Technology* 9(6) (2015b) 566-575.

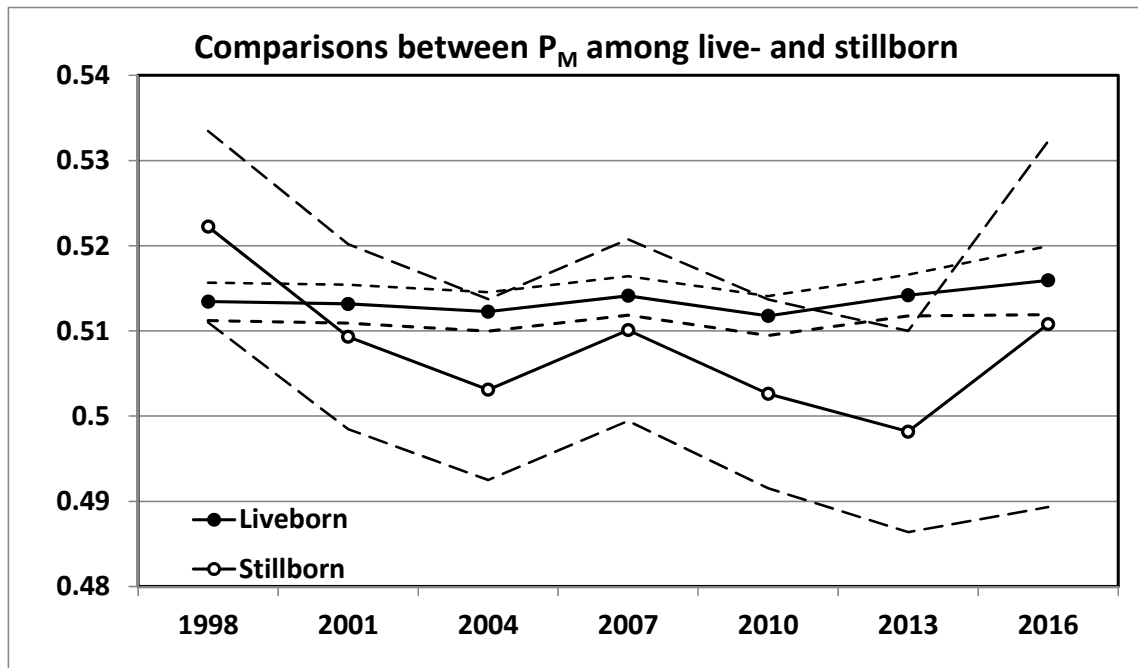


Figure 1. Temporal trends for proportions of males, P_M , among live- and stillborn children. The figure includes 95% confidence bands. The difference between live- and stillborn is non-significant. The study is based on Danish birth data for the period 1997-2016.

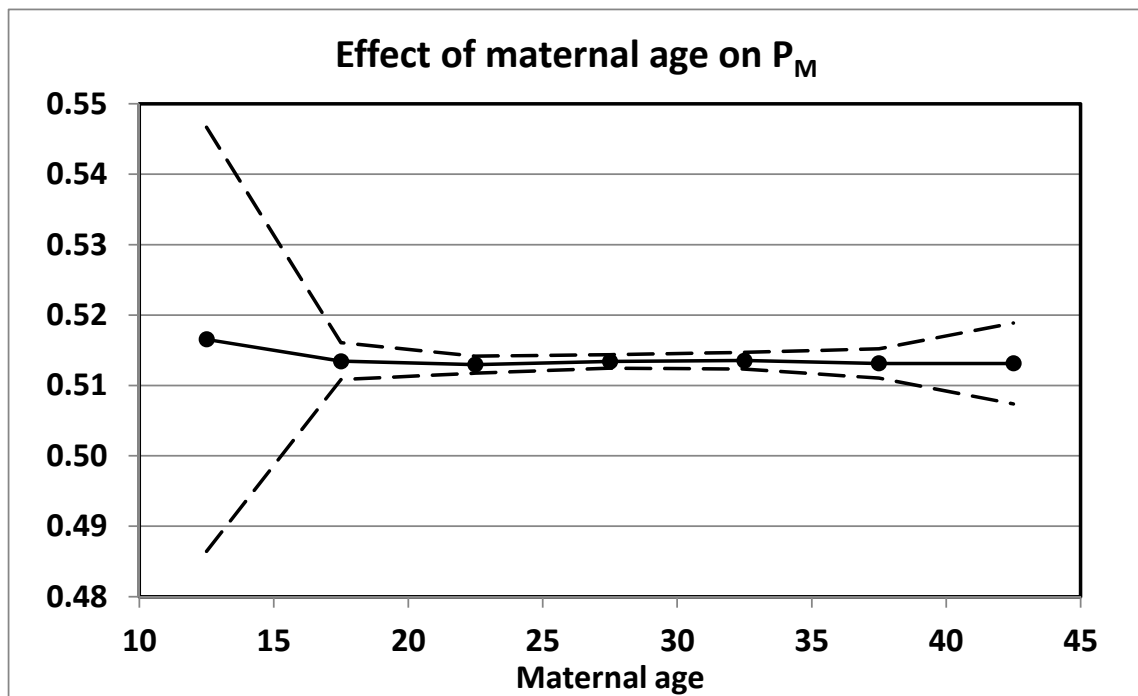


Figure 2. Effect of maternal age on the proportion of males, P_M . The figure includes a 95% confidence band. The analysis is based on Danish data for the period 1973-2016. The figure includes a 95% confidence band.

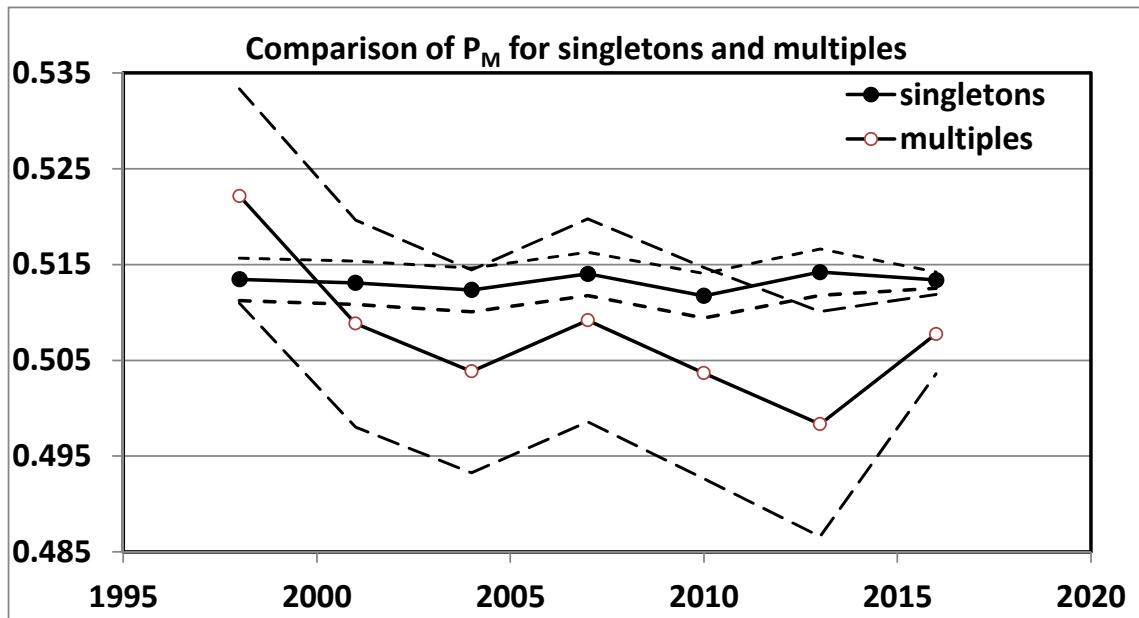
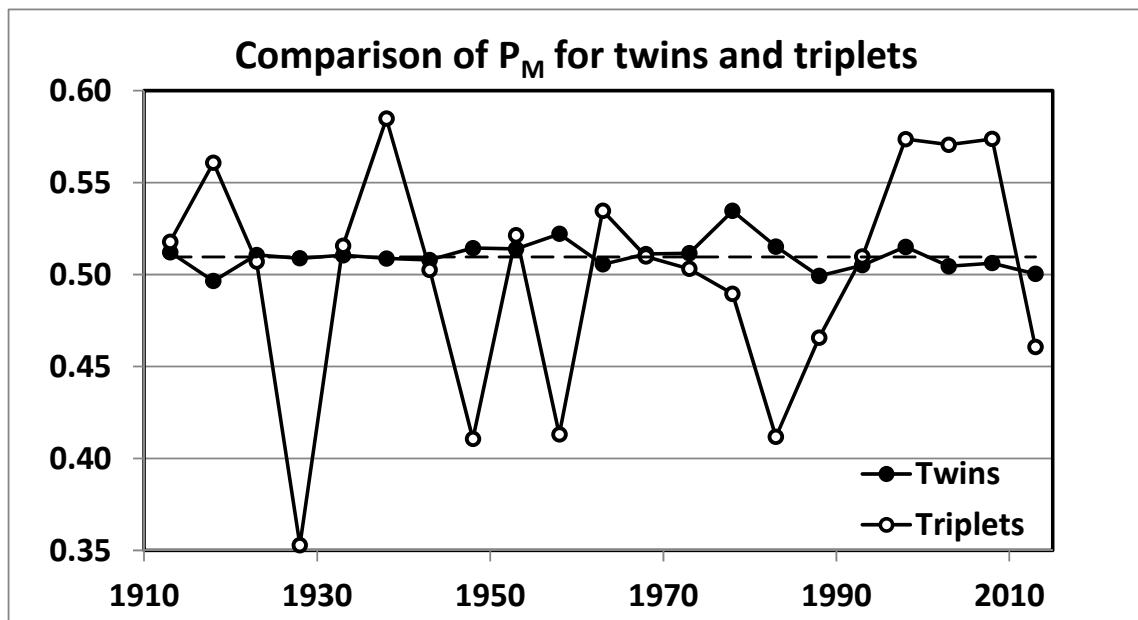


Figure 3. Temporal trends for the proportion of males, P_M among singletons and multiples. The analysis is based on Danish data for 1997-2016. The figure includes 95% confidence bands.



229

230 **Figure 4.** Temporal trends for the proportion of males, P_M among twins and triplets. The
 231 analysis is based on Danish data for 1911-2016. The figure includes the total mean for twins.