Twin Research (1999) 2, 22–29

http://www.stockton-press.co.uk/tr

Statistical analysis of the seasonal variation in the twinning rate

Johan Fellman and Aldur W Eriksson

Folkhälsan Institute of Genetics, Population Genetics Unit, Helsinki, Finland

There have been few secular analyses of the seasonal variation in human twinning and the results are conflicting. One reason for this is that the seasonal pattern of twinning varies in different populations and at different periods. Another reason is that the statistical methods used are different. The changing pattern of seasonal variation in twinning rates and total maternities in Denmark was traced for three periods (1855–69, 1870–94, and 1937–84). Two alternative methods of analysis are considered. The method of Walter and Elwood and a trigonometric regression model give closely similar results. The seasonal distribution of twin maternities for the periods in the 19th century showed highly significant departures. For both twin and general maternities, the main peaks can be seen from March to June and a local peak in September. During the spring–summer season the twinning rates were higher than the total birth rates, indicating a stronger seasonal variation for the twin maternities than for the general maternities. For 1937–84, there was a similar, but less accentuated, pattern. Studies of other populations are compared with the Danish results. The more accentuated seasonal variation of twinning in the past indicate that some factors in the past affected women during summer–autumn and around Christmas time, making them more fecund and particularly to be more prone to polyovulation and/or more able to complete a gestation with multiple embryos.

Keywords: multiple maternities, twinning rate, secular variation, deseasonality, legitimacy, fecundity, rural contra urban, Walter-Elwood model, trigonometric regression, Denmark

Introduction

There has been little agreement as to whether twin maternities show seasonal variation and especially about whether there is any universal pattern. The explanation may be that different populations have been studied for different periods and that the factors, both geophysical and socio-economic, that may influence the seasonal variation of twinning, differ in different populations. Furthermore, the statistics for twin maternities depend on different demographic and genetic factors, such as the age and parity distributions of the mothers and the population-specific twinning rates. These factors vary greatly between populations and over time.[1,2,3,4] In general, these factors should be considered in every study of twinning. Even if they do not markedly influence the pattern of seasonal variation they may influence the extent of seasonal variation when comparisons are made between different populations or of the same population at different times.

Another reason for the conflicting results may be that different statistical models and methods have been used. Every statistical model is based on some explicitly or implicitly assumed conditions and hence the chosen method identifies specific components of the seasonal pattern. Furthermore, the success of the statistical analysis depends on how well the assumptions on which the model is based correspond to the empirical data.

In the past, socio-economic status, general health, including physical condition, duration and intensity of daylight, supply of food, including famines, nutritional habits, work load etc. may have had a stronger influence on the seasonal variation of maternities, and especially of twin maternities. To-day, family planning influences the preferred time of birth. Consequently secular trends in the seasonal pattern are to be expected and it is valuable to make secular analyses of the seasonal variation of the twinning in different populations.

Methods

General

The history of the statistical analysis of time series started with decomposition of the time series into (multiplicative or additive) components measuring

Correspondence: Professor J Fellman, Folkhälsan Institute of Genetics, Population Genetics Unit, POB 211, FIN-00251 Helsinki, Finland. Tel: +358 9 61585554; Fax: +358 9 61588519; E-mail: fellman@shh.fi
Received 27 October 1998; revised 27 January 1999; accepted 29 January 1999

https://doi.org/10.1375/twin.2.1.22 Published online by Cambridge University Press

Table 1 Application of the Walter–Elwood model and the trigonometric regression model for testing seasonal variation in the total number of maternities and twin maternities in data and subdata for Denmark, 1855–69, total series for Denmark, 1870–94. For all series, the seasonal variation in the number of maternities is statistically highly significant. The angle θ* for the total number of births indicates a maximum mainly in March (60° < θ* < 90°) and for the twin maternities in April (90° < θ* < 120°). The lower values of the angle are estimated from our proposed regression analysis. For Denmark, 1937–84, the regression analysis gives the angle 136.5° indicating a maximum in May–June. As explained in the text, the Walter–Elwood method is not applicable in this case.

Data sets for Denmark	*All maternities*			*Twin maternities*				
		Test of the model		*Heterogeneity*			*Test of the model*	
	θ(°)	χ^{2a}	*P*	χ^{2b}	*P*	θ(°)	χ^{2a}	*P*
Total maternities, 1855–69	86.1 85.9	1429	< 0.001	38.0	< 0.001	95.0 95.5	22.8	< 0.001
Legit. maternities, 1855–69	87.6 87.2	1144	< 0.001	28.5	< 0.01	91.7 92.4	20.5	< 0.001
Illegit. maternities, 1855–69	78.0 79.0	323	< 0.001	26.6	< 0.01	123.0 121.0	2.8	> 0.05
Rural maternities, 1855–69	92.8 92.5	2780	< 0.001	30.1	< 0.001	89.6 90.1	7.2	< 0.05
Urban maternities, 1855–69	53.7 53.6	207	< 0.001	21.6	< 0.05	105.5 105.8	9.4	< 0.01
Total maternities, 1870–94	64.4 64.6	3336	< 0.001	41.2	< 0.001	112.3 113.2	20.3	< 0.001

[a]χ^2 values with 2 degrees of freedom.
[b]χ^2 values with 11 degrees of freedom.

time trends, seasonal variations, cyclical fluctuations and irregular (random) variations. In general, time series analysis is characterised by the assumption that the time series are auto(serially) correlated. Box and Jenkins[5] developed their general theory of how to model autocorrelation. Time series of twinning rates have been studied within the Box-Jenkins framework[6] but, in our opinion, the Box-Jenkins assumptions concerning autocorrelations are not well suited to the study of time series of twinning data. In our view, it is better to assume that the time series of twinning rates show time heterogeneity, that is, the observed data are independent although the parameters in their distributions depend on time.[7] Furthermore, we found that autocorrelation can be introduced into the model through explanatory variables showing autocorrelation. In fact, the statistics for twin maternities depend on a variety of demographic and genetic factors, such as the age and parity distributions of the mothers and the population-specific twinning rates. These factors vary greatly between populations and also over time.[1,3,4] However, none of the main factors can be considered to be autocorrelated.

The results obtained by statistical analysis depend on how the specific models used are built and tested. A simple model often gives statistically significant results, but the subsequent interpretation of the results is difficult. In contrast, a more sophisticated model for the seasonal variation may give statistically non-significant results, either because the seasonal variations are slight or because the goodness-of-fit of the proposed model is weak.

Several authors have proposed that monthly data should be pooled as quarterly data.[8,9] In our opinion, however, this should be done with great caution and tested empirically. Pooling must be based entirely on assumptions and a proposed model, but under no circumstances on the data obtained. Furthermore, such pooling may mask individual months with exceptional effects (eg September, see below).

Walter–Elwood model

The model proposed by Walter and Elwood[10] has been successfully used in studies of seasonal variations. The model assumes a cyclic trend of epidemiological events and is a modification of a method proposed by Edwards.[11] The seasonal pattern is described by polar coordinates (r,θ), where the length of the radius vector r is the square root of the monthly values and the angle θ represents the month. Thus, every month corresponds to an angle of approximately 30° (the whole year corresponding to 360°). A geometric picture of the model is given by Walter and Elwood.[10]

In their method, a central role is played by the 'population at risk'. For example, if we study the seasonal variation in twin maternities or in the occurrence of an innate disease, the population at risk is the total number of confinements. Hence, the number of twin maternities in a given month must be compared with the monthly number of all maternities. In a study of the seasonal variation in the number of births the population at risk is the product of the size of the population times the length of the month. For short periods, the population can be assumed to be constant and therefore the weights are proportional to the lengths of the months.

We have developed the method of Walter and Elwood in the following way. Using their notations, we have redefined their angle θ^*. Their basic definition of θ^* was

$$\tan\theta^* = \frac{\bar{y} - \mu_{\bar{y}}}{\bar{x} - \mu_{\bar{x}}} \quad [1]$$

where $(\mu_{\bar{x}}, \mu_{\bar{y}})$ is the theoretical centroid and $(\bar{x}, \bar{y})$ is the observed centroid of the data.

If $\tan\theta^* > 0$, Walter and Elwood chose an acute positive angle θ^* ($0° \le \theta^* \le 90°$) and if $\tan\theta^* < 0$ they chose an acute negative angle θ^* ($-90° \le \theta^* \le 0°$). However, the angle θ^* can be given a more illustrative interpretation if it is defined in the following way:

If $\tan\theta^* > 0$ and $\bar{x} - \mu_{\bar{x}} > 0$ then $0° \le \theta^* \le 90°$

If $\tan\theta^* < 0$ and $\bar{x} - \mu_{\bar{x}} < 0$ then $90° \le \theta^* \le 180°$

If $\tan\theta^* > 0$ and $\bar{x} - \mu_{\bar{x}} < 0$ then $180° \le \theta^* \le 270°$

If $\tan\theta^* < 0$ and $\bar{x} - \mu_{\bar{x}} > 0$ then $-90° \le \theta^* \le 0°$ or $270° \le \theta^* \le 360°$ [2]

Combining the Walter and Elwood definition (1) and our definition (2), we obtain an angle θ^* which points from the theoretical centroid $(\mu_{\bar{x}}, \mu_{\bar{y}})$ towards the observed centroid $(\bar{x}, \bar{y})$. It is in this direction that the months with high rates can be found. Hence, the angle θ^* is associated with the months with high values.

St Leger[12] also presented a modification of the Edwards test. He assumed that the pattern of seasonal variation is sine-shaped and he estimated a sine curve by a maximum likelihood (ML) approach. As an alternative method, we consider a regression approach. Consider the model

$E(Y_i) = A + R\sin(t_i + \alpha) = A + R\sin\alpha\cos t_i + R\cos\alpha\sin t_i = A + B_1\cos t_i + B_2\sin t_i$,

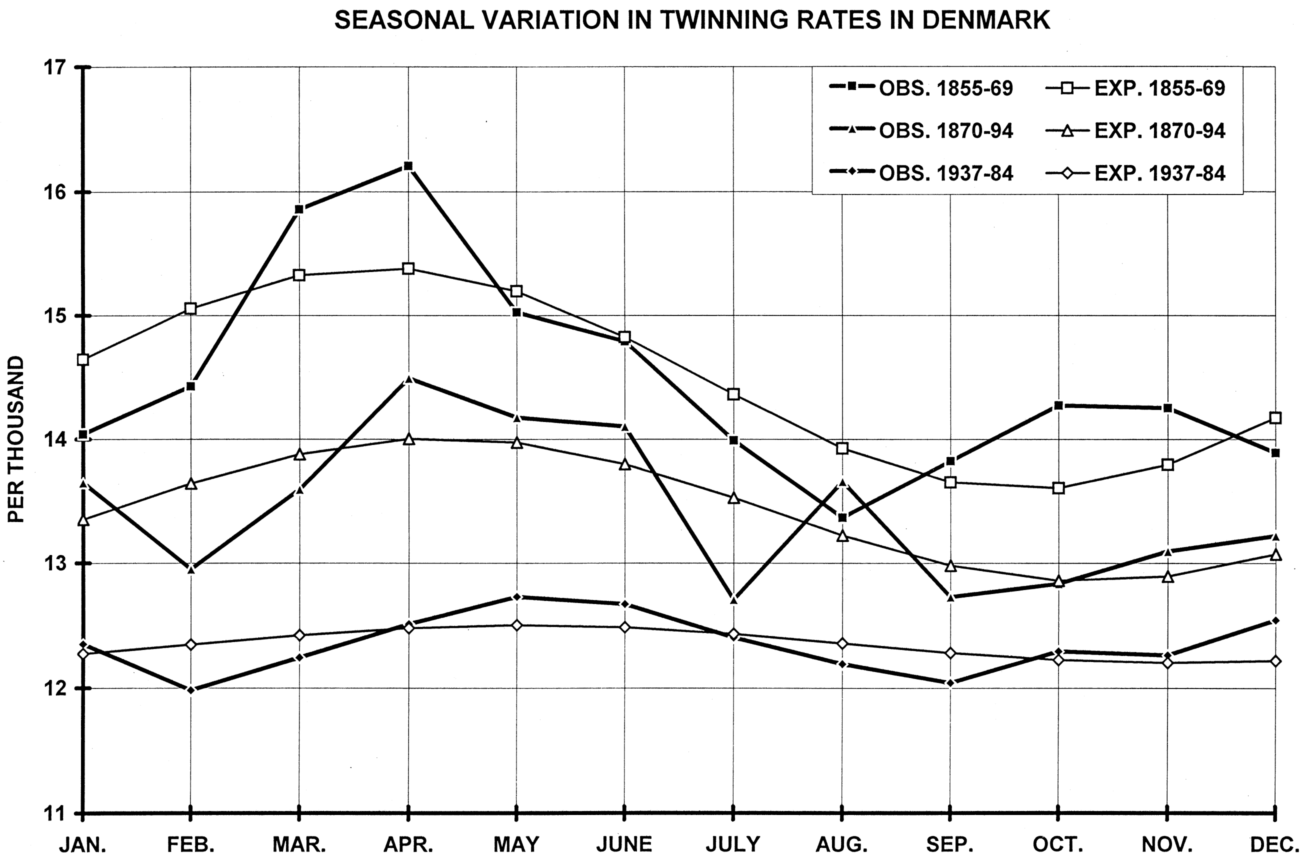

Figure 1 Seasonal variation in the twinning rate in Denmark for the periods 1855–69 ($\theta^* = 95.5°$), 1870–94 ($\theta^* = 113.2°$) and 1937–84 ($\theta^* = 136.3°$). We observe a shift of the maximum from April to May. The expected curve is the estimated regression model. Both the twinning rate and its seasonal variation show absolute decreases. The observed total twinning rates for the three periods are 14.5, 13.4 and 12.4 per thousand confinements. The respective SDs measuring the strength of the seasonal variation are 0.84, 0.60 and 0.23

https://doi.org/10.1375/twin.2.1.22 Published online by Cambridge University Press

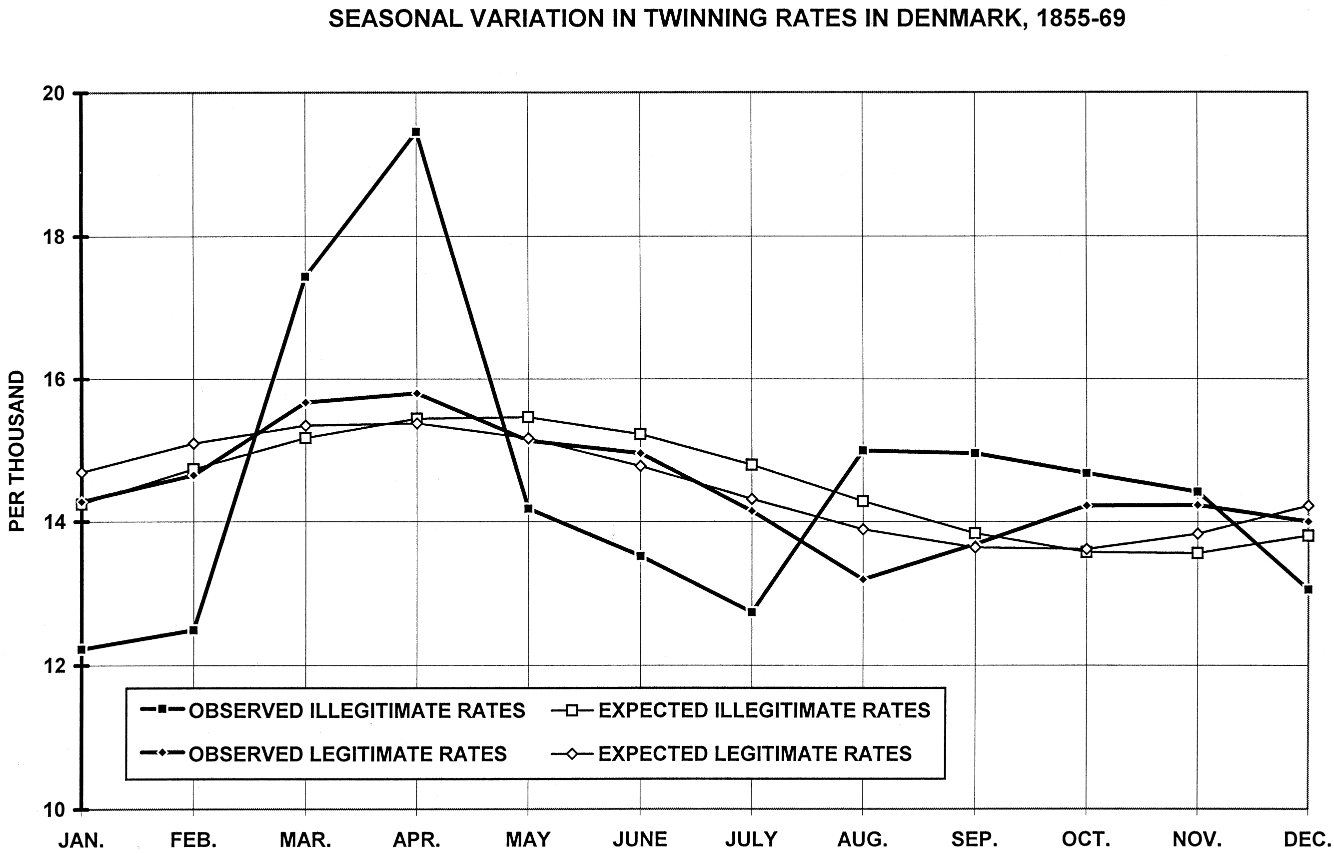

Figure 2 Comparison of the seasonal patterns of legitimate ($\theta^* = 92.4°$) and illegitimate ($\theta^* = 121.0°$) twin maternities in Denmark, 1855–69

where Y_i is the observed rate, and t is an angle representing the time (in months) as described above, and $B_1 = R\sin\alpha$ and $B_2 = R\cos\alpha$.

The parameters (A, B_1, B_2) are estimated by the ordinary least squares (OLS) method for monthly data. From these estimates, it is possible to estimate the intercept $\hat{A}$, the amplitude $\hat{R} = \sqrt{\hat{B}_1^2 + \hat{B}_2^2}$ and the angle (month) for the maximum. The maximum is obtained for $\hat{t} + \hat{\alpha} = 90°$, where $\tan(\hat{\alpha}) = \frac{\hat{B}_1}{\hat{B}_2}$. The angles $\hat{\alpha}$ and $\hat{t} = 90° - \hat{\alpha}$ are chosen according to the same rules as θ^* in formula (2). Hence, the angle $\hat{t} = 90° - \hat{\alpha}$ if $\hat{\alpha} \leq 90°$ or $\hat{t} = 450° - \hat{\alpha}$ if $\hat{\alpha} > 90°$ can be compared with the estimated angle θ^* in the Walter–Elwood model. Our new proposed method has the advantage that it can be applied in situations, in which only the monthly twinning rates are available. This is not the case with the Walter–Elwood method or the St Leger method.

Statistical tests

In different studies of seasonal variation, several alternative tests have been proposed. Seasonal heterogeneity can be tested by a standard χ^2 test with 11 degrees of freedom ($\chi^2_{crit} = 19.68$). This test is distribution free but significance depends on marked monthly differences and is difficult to interpret.

Freedman[13] introduces a modified Kolmogorov–Smirnov test. In short, he considers the difference between the cumulative proportion of days and the cumulative proportion of events (maternities, births etc). If there is no seasonal variation, these two cumulative proportions should increase together. This test identifies any discrepancies and no assumption of a sine-shaped pattern is needed.

An easily performed, model-free, relative measure of extent of the seasonal variation is the standard deviation (SD) calculated from the monthly index data, as used, for instance, by Rosenberg.[14] However, this measure cannot be used for analyses of specific seasonal patterns. Anderson and Silver[15] used the coefficient of variation (CV) as a measure of the heterogeneity in the monthly number of births. Compared with the SD, their measure has the advantage that it is independent of the scale of the data.

The cyclic trend can be tested by a χ^2 test proposed by Walter and Elwood, with two degrees of freedom

https://doi.org/10.1375/twin.2.1.22 Published online by Cambridge University Press

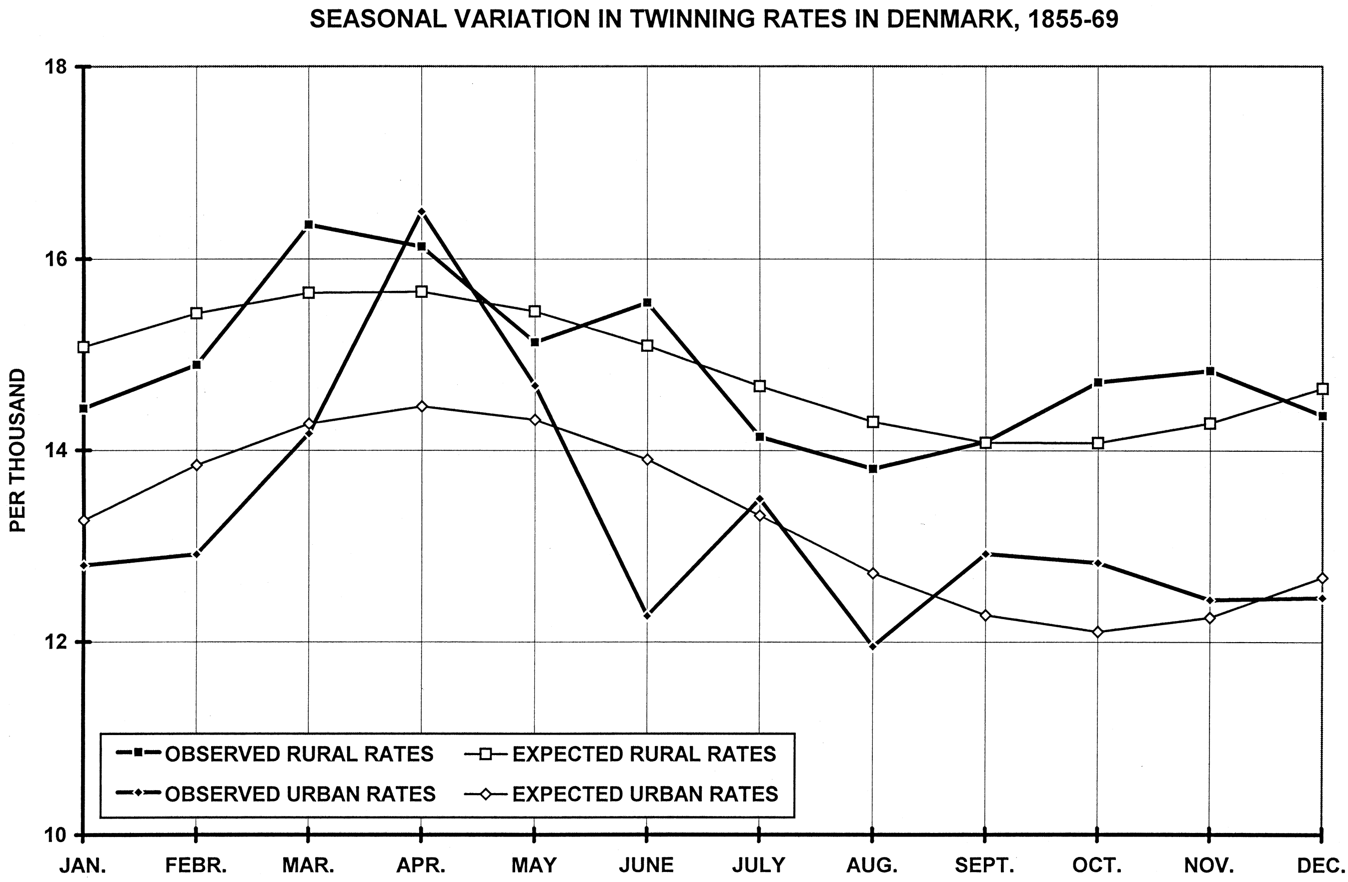

Figure 3 Comparison of the seasonal patterns of rural (θ^* = 90.1°) and urban (θ^* = 105.8°) twin maternities in Denmark, 1855–69

Table 2 Results of the Walter–Elwood model and trigonometric regression first and second lines in each group) for total and twin maternities in data for Switzerland, 1876–90, for England and Wales, 1952–59 and 1963–75, for north–eastern Scotland and for Northern Ireland, 1975–79

Data set	*Test of total maternities corrected for the length of the month* θ	*Highest monthly* rate	*Test of twin maternities corrected for total maternities* θ	*Highest monthly* rate	χ^{2a}
Switzerland, 1876–90	110.8	April	107.3	April	2.33
	110.7		107.7		
England and Wales, 1952–59 and	111.1	April	– 23.9	Dec.	**39.34**
1963–75	110.8		– 23.6		
NE Scotland, 1975–79	170.7	June	32.5	Feb.	1.17
	170.6		33.1		
N Ireland, 1975–79	145.5	May	72.1	March	1.81
	145.1		71.9		

[a]χ^2 values with 2 degrees of freedom.

(χ^2_{crit} = 5.99). The null hypothesis assumes no seasonal variation and the alternative hypothesis is a sinusoidal model. Non-significant χ^2 values may be caused by slight seasonal variation or by discrepancies between the assumed model and the data. In general, χ^2 tests presuppose absolute values, and therefore this test is worthless if the only data available are monthly indices or rates. However, under such circumstances, the χ^2 statistic can be used as a measure of the relative extent of the seasonal variation pattern. St Leger (1976) presented a modification of the Edwards test. His test was based on his ML method but it has the same shortcomings as the Walter–Elwood test. These tests and alternatives were discussed by Reijneveld.[16]

In our opinion, the sinusoidal tests should be combined with more general tests. In our studies, the data obtained show strong seasonal variation but the seasonal pattern often differs from a sine curve. We observe in our series a main maximum (usually in spring) and an isolated peak in September. This indicates that the Walter–Elwood model does not give the best fit. Statistical models and

https://doi.org/10.1375/twin.2.1.22 Published online by Cambridge University Press

corresponding tests for the analysis of seasonal variation are presented in greater detail elsewhere.[17]

Results and discussion

General

The fact that the data available are for different populations and for different time periods complicates the study of the secular trends of the seasonal variation of twinning. It is (almost) impossible to find long time series for different populations containing sufficient amounts of comparable information. As a rule, one has to deal with separate sets of data from different time periods and compare the results obtained in order to gain an impression of any secular change in the seasonal variation of twinning. This temporal and regional sparcity of available data also weakens any attempt at a systematic study of geographical and climatological effects on the seasonal variation. Therefore, in this study, we will use the model of Walter–Elwood and our regression model and concentrate our study mainly on the seasonal pattern of twin maternities.

A goodness-of-fit test based on a specific model gives the answer to how well the data can be explained by this model. Within the framework of this study, it does not give a straight answer to the question of whether there is, in general, any seasonal heterogeneity. Therefore, we first test the heterogeneity by the standard χ^2 test (with 11 degrees of freedom). We then perform the goodness-of-fit test based on the Walter-Elwood model. The results obtained with the Walter–Elwood model are compared with the results of the regression model. In the figures, the expected curves are the estimated regression models.

Available data

Our analysis is based mainly on data from Denmark for the periods 1855–1869,[18] 1855–94[19] and 1937–84.[20] The data given by Weinberg are divided into two periods, 1855–69 and 1870–94. Consequently, we have twin maternity data for three separate periods, 1855–69, 1870–94 and 1937–84. For the period 1855–69, Neefe[18] gives subdata for legitimate maternities, illegitimate (extramarital) maternities, urban maternities and rural maternities. For the other periods, data are given only for the total population. Our findings for Denmark are compared with studies of twin data for Switzerland, 1876–90,[19] for England and Wales, 1952–59, 1963–75,[21] and for north-east Scotland and for Northern Ireland, 1975–79.[22]

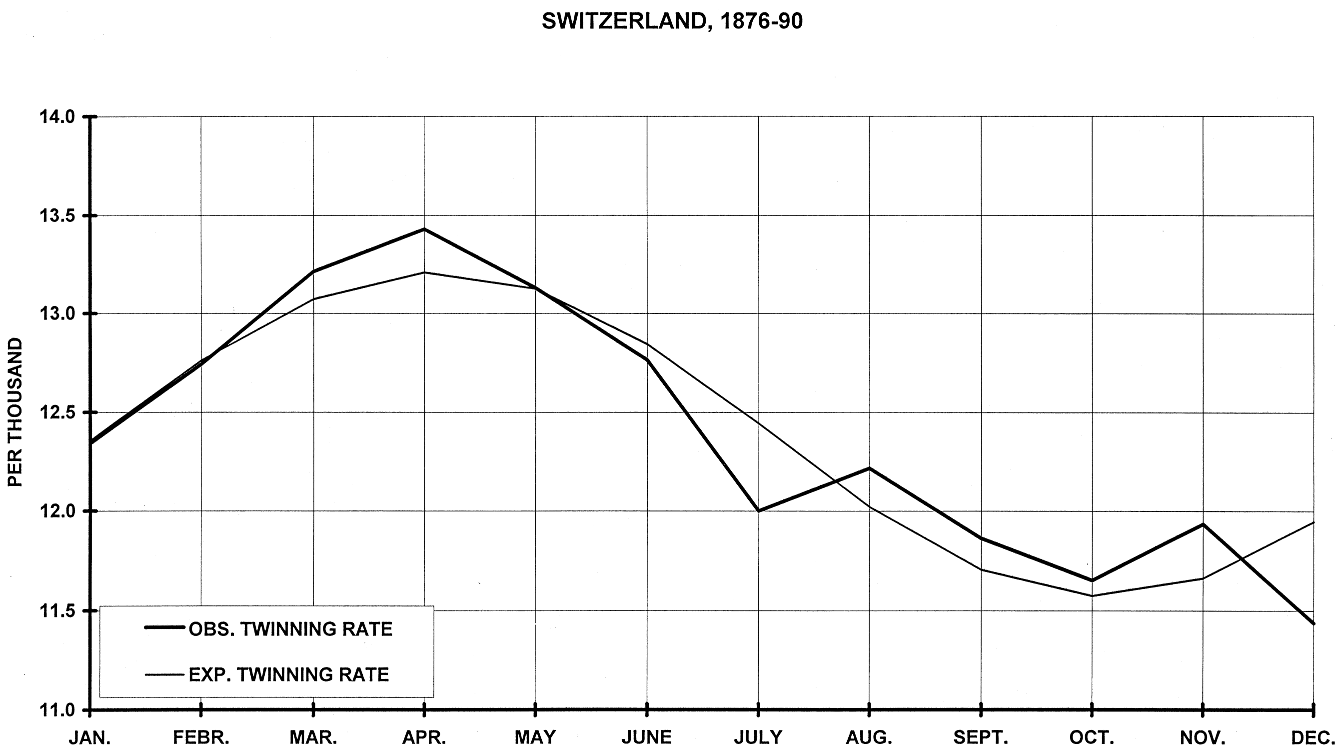

Figure 4 Seasonal variation of observed and expected twin maternities in Switzerland, 1876–90 ($\theta^* = 107.7°$)

https://doi.org/10.1375/twin.2.1.22 Published online by Cambridge University Press

Denmark

Our results for Denmark are summarised in Table 1. In the first line of each group we give the results according to the Walter–Elwood model and in the second line those of our regression analysis. For all series, the monthly number of maternities show statistically strongly significant seasonal heterogeneity. The application of the Walter–Elwood model to the total number of maternities also gives statistically significant sinusoidal departures, maximum birth rates occurring in March and minimum rates in autumn. Compared with our studies of births in Sweden, the results are quite different. During the period 1851–1900 the maximum birth rate in Sweden moved from January to February ($10.6 \leq \theta^* \leq 38.6$). Not until this century did the maximum in Sweden move to April.[23]

Secular variations in the seasonal pattern In Figure 1 we compare the seasonal variations in the total twinning rate in Denmark for the periods 1855–69, 1870–94 and 1937–84. The twinning rate shows an absolute decrease. The same can be said about the strength of the seasonal variation. For the time periods 1855–69, 1870–94 and 1937–84, the corresponding observed total twinning rates are 14.5, 13.4 and 12.4 per thousand, respectively. If we use the SD as a measure of the extent of the seasonal variation in the twinning rates, we obtain values of 0.84, 0.60 and 0.23, respectively. At the same time there is a shift in the maximum from April to May.

Bonnelykke *et al*[20] found a marked decline in the twinning rate during the period 1937–84. They eliminated this trend with a polynomial of the 5th degree. They then performed a thorough study of the seasonal variation during the period 1937–84, and noted a slight seasonal variation in the data. Tests with the harmonic sinusoidal model showed no significant differences but a statistically significant polynomial model for the intra-year variation was found.

Seasonal variation according to marital status and residence Legitimate maternities and illegitimate maternities are compared in Figure 2, rural maternities and urban maternities in Figure 3. We observe a general pattern, the twin maternities showing stronger seasonal fluctuations than the total maternities, in all subsamples. The spring maximum is most marked in the data for illegitimate maternities and urban maternities.

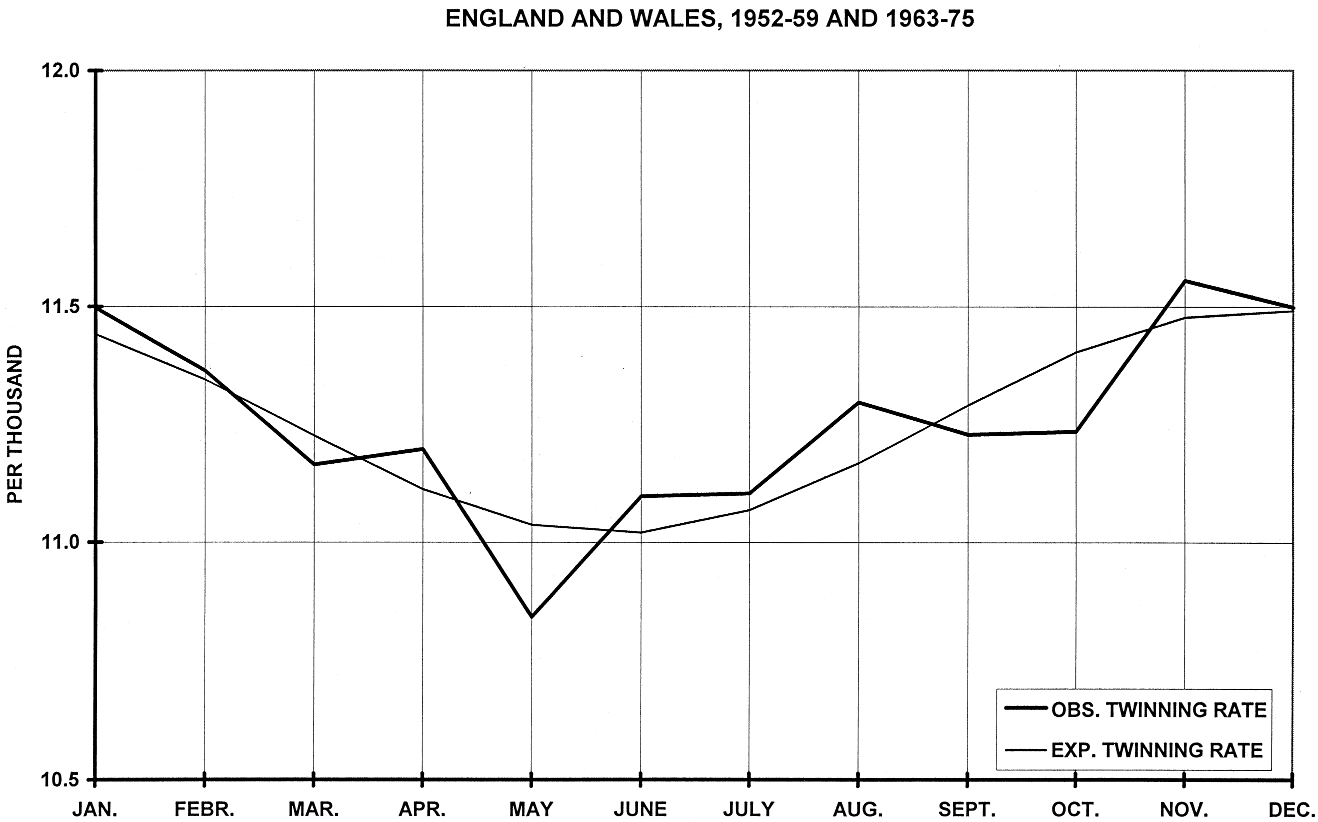

Figure 5 Seasonal variation of observed and expected twin maternities in England and Wales, 1952–59 and 1963–75 ($\theta^* = 336.4°$). Observe that the θ^* value indicates a maximum in December

https://doi.org/10.1375/twin.2.1.22 Published online by Cambridge University Press

Only for the series of extramarital maternities is the sinusoidal pattern not significant. The values obtained for the angle θ^* indicate that the twinning rates for the illegitimate maternities, compared with the total number of maternities, reach extreme values in March and April. These conceptions had taken place during the summer months June and July and also the first half of August if we allow for the fact that the average gestation time is about 3 weeks shorter for twin than for singleton births. The unexpected result that the illegitimate births show strong seasonal fluctuations (Figure 2) but that these are not significant is partly explained by the fact that the data set is rather small and partly by the fact that the data do not fit the Walter–Elwood model. The same also holds in part for the urban births (Figure 3).

Comparisons with other populations

We have applied the models to the series of total births of some neighbouring nations of Denmark. The results are given in Table 2. It is notable that the data for Switzerland and for England and Wales indicate a maximum for the total number of births in April, but for Northern Ireland in May and for north-east Scotland in June.

The twin maternities in Switzerland show a seasonal variation similar to Denmark but this is not significant (Figure 4). The twinning data for England and Wales show statistically highly significant seasonal variation but with a maximum peak in December (Figure 5). No significant pattern can be found in the data from north-east Scotland (maximum in February) or Northern Ireland (maximum in March).

In conclusion, our results show seasonal variations in twinning rates, which were more accentuated in the past. The marked changes in the seasonal pattern observed in all births, especially in Sweden, between 1841 and 1991,[23] is not seen in this analysis of twinning data from Denmark. Studies of the seasonal variation in twinning demand large and comparable series. This fact and the discernible deseasonality (in Denmark) may at least partly explain the conflicting results earlier reported.

Acknowledgements

The authors would like to thank the foundation ‘Liv och Hälsa’ for supporting the project. This article is based on a paper presented at the Ninth International Congress on Twin Studies, Helsinki, 4–6 June 1998.

References

1 Eriksson AW. Variations in the human twinning rate. *Acta Genet (Basel)* 1962; **12**: 242–250.
2 Eriksson AW. Human twinning in and around the Åland Islands. *Commentationes Biologicae (Diss)* 1973; **64**: 1–159.
3 Eriksson AW, Eskola MR, Fellman JO. Retrospective studies on the twinning rate in Scandinavia. *Acta Genet Med Gemellol (Roma)* 1974; **25**: 29–35.
4 Eriksson AW, Fellman J. Differences in the twinning trends between Finns and Swedes. *Am J Hum Genet* 1973; **25**: 141–151.
5 Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control*. Rev. edn. Holden-Day: San Francisco, 1976.
6 Philippe P. Box-Jenkins modelling of chronological series of twin births and single births. *Hum Biol* 1991; **63**: 367–387.
7 Spanos A. *Statistical Foundations of Econometric Modelling*. Cambridge University Press: Cambridge, 1995, ch. 17.
8 Nonaka K, Miura T. Incidence of twinning in London from 1581 to 1760. *Acta Genet Med Gemellol (Roma)* 1987; **36**: 365–372.
9 Nonaka K, Miura T. Principles in methods of epidemiological studies on birth seasonality. In: T. Miura (ed.). Seasonality of Birth, *Progress in Biometeorology* **6**. SPB Academic Publishing: The Hague, 1987, pp 13–24.
10 Walter SD, Elwood JM. A test for seasonality with a variable population at risk. *Brit J Prev Soc Med* 1975; **29**: 18–21.
11 Edwards JH. The recognition and estimation of cyclic trends. *Ann Hum Genet* 1961; **25**: 83–87.
12 St Leger AS. Comparison of two tests for seasonality in epidemiological data. *Appl Statist* 1976; **25**: 280–286.
13 Freedman LS. The use of Kolmogorov–Smirnov type statistic in testing hypotheses about seasonal variation. *J Epidemiol Commun Health* 1979; **33**: 223–228.
14 Rosenberg HM. *Seasonal Variation in Births. United States, 1933–63*. Public Health Service Publication 1000, Series 21, Number 9. US Department of Health, Education and Welfare: Washington DC, 1966.
15 Anderson BA, Silver BD. The effects of the registration system on the seasonality of births: The case of the Soviet Union. *Pop Studies* 1988; **42**: 303–320.
16 Reijneveld SA. The choice of a statistic for testing hypotheses regarding seasonality. *Am J Phys Anthrop* 1990; **83**: 181–184.
17 Fellman J, Eriksson AW. Statistical analysis of the seasonal variation of demographic data. (submitted).
18 Neefe M. Zur Statistik der Mehrgeburten. *Jahrb Nationalökon Stat* **28**: 168–194.
19 Weinberg W. Beiträge zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen. *Arch Ges Physiol* 1901; **88**: 346–430.
20 Bonnelykke B, Søgaard J, Nielsen J. Seasonality in twin birth rates, Denmark, 1936–84. *J Epidem Commun Health* 1987; **41**: 338–343.
21 James WH. Seasonality in twin and triplet births. *Ann Hum Biol* 1980; **7**: 163–175.
22 Little J, Elwood JM. Problems in the assessment of seasonality in twinning rates with reference to twinning in North-East Scotland and Northern Ireland, 1975–1979. In: T. Miura (ed.). Seasonal Effects on Reproduction, Infection and Psychoses, vol. 5, *Progress in Biometeorology*. SPB Academic Publishing: The Hague, 1987, pp 61–72.
23 Fellman J, Eriksson AW. Secular changes in the seasonality pattern of births in Nordic Countries. *Perspect Hum Biol* 1999; **4** (in press).