

Twinning Rates in Isolates

Johan Fellman

Hanken School of Economics, Helsinki, Finland

The aim of this study was to investigate the twinning rates (TWRs) in isolates relative to the TWRs in the surrounding populations. It is not uncommon that the TWR shows extreme values (high or low rates) within isolated subpopulations. Starting from the isolated populations of the Åland Islands in Finland (high rates), we enlarged our studies to other isolated subpopulations in other countries: the island of Gotland (high rates), the county of Älvsborg located in the southwestern part of Sweden (low rates), and mountain villages in Norway. In our statistical analyses, we paid special attention to the robustness of the variance formula of the TWR and to alternative confidence intervals for the TWR. Particularly, we show how to obtain the most precise confidence intervals for the twinning rates. These statistical methods are crucial when the extreme TWRs within subpopulations are compared with the TWRs within the general population. One must decide whether the differences are real or caused by random fluctuations within the small isolates.

■ **Keywords:** confidence interval, counties, Norway, Sweden, Gotland, Älvsborg, Finland, Åland Islands

This study aimed to investigate the twinning rates (TWRs) in isolates compared with the rates in the surrounding majority populations. The TWR often shows extreme values (high or low rates) within isolated subpopulations. In this study, we consider isolates in Finland (Åland Islands), Sweden (island of Gotland and county of Älvsborg), and Norway (mountain villages). The Finnish study is based on Eriksson (1973), the Swedish study on Fellman and Eriksson (2003, 2005b, 2009), and the Norwegian study on Bonnevie and Sverdrup (1926). In our statistical analyses, we paid special attention to: (a) the robustness of the variance formula of the TWR, and (b) the most precise confidence intervals for the TWRs.

Fellman and Eriksson (2004, 2006) and Fellman (2013) have considered these aspects of the statistical analyses of the TWRs more thoroughly. These statistical methods are crucial when the extreme TWRs within often small subpopulations are compared with the TWRs within the general population. One must decide whether the differences are real or caused by random fluctuations within the small isolates.

Methods

Variance of Observed Proportions

In population studies of the relative frequency of twin maternities, the accepted variance formula for the observed

rates is

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}, \quad (1)$$

where p is the theoretical probability of success.

Consider n repeated Bernoulli trials, that is, trials where the probability of success in a specific trial is independent of earlier outcomes. Let S be the total number of successes (e.g., twin maternities) and let $\hat{p} = \frac{S}{n}$ be the observed proportion of successes. For a large n , both S and \hat{p} are asymptotically normal. If the assumption holds that (a) the repetitions are independent and (b) the probability of success, p , is constant during the repetitions, then the variance formula (1) for the observed proportion \hat{p} holds.

Usually there is no discussion about the premises, and so we can use formula (1). In some situations, however, such a discussion may be necessary. Let us consider the outcome of the maternities of a certain group of mothers. The maternities (the mothers) are the repetitions, and success is the birth of a twin set. If we consider different mothers,

RECEIVED 30 May 2016; ACCEPTED 24 August 2016. First published online 27 September 2016.

ADDRESS FOR CORRESPONDENCE: Professor Johan Fellman, Hanken School of Economics, POB 479, FI-00101 Helsinki, Finland. E-mail: fellman@hanken.fi

then we can assume independent repetitions. However, the constancy of the binomial proportion is more difficult to accept. We know that the probability of a twin maternity varies greatly, depending on several factors, particularly the individual factors of maternal age, parity and ethnicity. Hence, the variance formula (1) for the total TWR cannot hold exactly. Does it hold approximately and with what accuracy? Cramér (1951, p. 206) has studied the effect of defects in assumption (b) on the accuracy of the variance formula.

In order to generate classes that are as homogeneous as possible, assume that the mothers are grouped in classes according to presumptive influential factors. The statistical analysis performed is based on the crucial assumption that the grouping is chosen before the outcome is known. Let the number of mothers in group number r be n_r ($r = 1, \dots, R$) and the total number of mothers be n , that is, $n = \sum_r n_r$. Furthermore, let S_r ($r = 1, \dots, R$) be the observed number, and $\hat{p}_r = \frac{S_r}{n_r}$ be the observed TWR within group number r and let $S = \sum_r S_r$. Assuming constant probability within the groups, the theoretical probability of twin maternities in the total population is $p = \frac{1}{n} \sum_r n_r p_r$, where p_r is the group-specific probability. The total observed relative frequency of successes is

$$\hat{p} = \frac{1}{n} \sum_r n_r \hat{p}_r = \frac{1}{n} \sum_r S_r = \frac{S}{n}. \tag{2}$$

The expectation is $E(\hat{p}) = \frac{1}{n} \sum_r n_r E(\hat{p}_r) = \frac{1}{n} \sum_r n_r p_r = p$, and \hat{p} is an unbiased estimator of p . Note that the estimator \hat{p} in (2) is the same, irrespective of whether we consider grouping or not. If we assume a constant probability within the classes but not between them, we obtain (Cramér, 1951; Fellman & Eriksson, 2004)

$$\begin{aligned} \text{Var}(\hat{p}) &= \frac{1}{n^2} \sum_r n_r^2 \text{Var}(\hat{p}_r) = \frac{1}{n^2} \sum_r n_r^2 \frac{p_r(1-p_r)}{n_r} \\ &= \frac{p(1-p)}{n} - \frac{1}{n^2} \sum_r n_r (p_r - p)^2. \end{aligned} \tag{3}$$

This fundamental result can also be written

$$\frac{p(1-p)}{n} = \frac{1}{n^2} \sum_r n_r p_r (1-p_r) + \frac{1}{n^2} \sum_r n_r (p_r - p)^2. \tag{4}$$

The left-hand-side in (4) is the total variation, giving the variance when one ignores any grouping. The first part on the right-hand side is the variation within the groups and, according to (3), it is $\text{Var}(\hat{p})$ when the grouping is considered. The second part is the variation between the groups. If, and only if, the rate is the same within all the groups, that is, $p_r = p$ for all $r = 1, \dots, R$, the second term is equal

to zero and all the variation is intra-group variation. In this case, the grouping factors have no influence on the incidence of twinning and the standard formula (1) for the variance holds. The first term on the right-hand side of (4) is zero if, and only if, $p_r = 1$ or $p_r = 0$. This means that the set of mothers is divided into groups consisting entirely of mothers with or without twin maternities. Such a grouping in surely homogeneous classes prior to the outcomes is possible only if every mother forms her own group. In this case, the total variation consists of inter-group variation and formula (1) holds.

The results obtained have an interesting interpretation. Without any grouping, the variance formula (1) holds. A grouping giving additional information about the variation of the probability reduces the variance. The reduction from (1) to $\text{Var}(\hat{p}) = \frac{1}{n^2} \sum_r n_r p_r (1-p_r)$ indicates the effect of grouping. If the variation in the p_r s is large, then the grouping factors (race, maternal age, parity, marital status, time, rural or urban population, season, etc.) are informative with respect to the TWR. On the other hand, if the variation in the p_r s is very small then the grouping factors are of small informative value, and consequently, the reduction from (1) to (3) can be ignored. Furthermore, the overestimation indicates that all statistical tests give excessively low test values, that is, the tests are conservative.

According to Fellman and Eriksson (2004), both the classical variance formula and the variance formula (3) are over-estimates as long as the groups are still heterogeneous. This follows from the fact that the formulae are based on the assumption that the probability of a twin maternity is constant within every group. Within these groups, additional factors, not considered, may cause unidentifiable heterogeneity. Such variations may be caused by inter-individual differences in the probabilities. Fellman and Eriksson (2004) observed that the corrections are minute, but in studies where one considers small series, great differences between groups, or large numbers of groups (cf. grouping according to individual mothers, discussed above), the corrections may be notable. The formulae derived above are the only ones available for correcting the variance of group heterogeneity in the probabilities for twin maternities. It is very common that registered twinning data are pooled, and later it is impossible to split the data into more homogeneous groups.

Confidence Interval for an Unknown Proportion

Now, we consider confidence intervals (CIs) for an unknown proportion p . The ML estimator of p is $\hat{p} = \frac{S}{n}$, where S is the number of successes (in our study, the number of twin maternities) and n is the total number of observations (all maternities). It is a well-known fact that \hat{p} has the mean p , is asymptotically normal, and, if we consider ungrouped data or data for a specific group, the variance $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$ is given in (1). The standard method, at least in applied studies, is that the variance formula is

estimated so that p is replaced by \hat{p} , resulting in the approximate test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \tag{5}$$

Based on this formula, the 100(1- α)% CI, the so-called Wald's CI, is

$$\left(\hat{p} - z_{1/2\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1/2\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right), \tag{6}$$

where $z_{1/2\alpha}$ is the (1- $1/2\alpha$) quantile for the standardized normal distribution.

Brown et al. (2001; 2002) gave a thorough presentation of the problems concerning CIs for unknown proportions. They showed that the standard method (6) in universal use is riddled with problems. The actual coverage probability for p can differ significantly from the nominal confidence level at realistic and even larger than realistic sample sizes. The error comes from two sources: discreteness and skewness ($p \neq 0.5$) in the underlying distribution. For a two-sided interval, the rounding error due to discreteness is asymptotically dominant, being of the order $n^{-1/2}$. The error due to skewness is of the order n^{-1} , but still important for even moderately large n . For one-sided intervals, the error caused by skewness can be larger than the rounding error. Brown et al. (2002) applied Edgewood expansions and found that two-term expansions fitted very well. This finding is a consequence of the fact that the first two terms in the Edgewood expansions are of the orders $n^{-1/2}$ and n^{-1} , respectively, being of the same order as the errors presented above.

The standard CI is centered at $\hat{p} = \frac{S}{n}$. Although \hat{p} is an unbiased ML estimator of p , as a center of a CI, it causes a systematic negative bias in the coverage. Brown et al. (2001) stressed that even in cases where the textbooks indicate that the formula (6) is safe, the coverage probability may differ markedly from the expected. As alternatives, they proposed the Wilson (1927) or the Agresti–Coull CI (Agresti & Coull, 1998). The alternative methods are briefly presented below.

The Wilson Interval

Using the notations introduced above, we consider the exact test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \tag{7}$$

which, on theoretical grounds, can be considered more reliable than the approximate one in (5). From (7) we obtain, after some calculations, the 100(1- α)% *Wilson confidence*

interval

$$\left(\frac{n\hat{p} + 1/2z_{1/2\alpha}^2}{n + z_{1/2\alpha}^2} - \frac{z_{1/2\alpha}\sqrt{n}}{n + z_{1/2\alpha}^2} \sqrt{\hat{p}(1-\hat{p}) + \frac{z_{1/2\alpha}^2}{4n}}, \right. \\ \left. \frac{n\hat{p} + 1/2z_{1/2\alpha}^2}{n + z_{1/2\alpha}^2} + \frac{z_{1/2\alpha}\sqrt{n}}{n + z_{1/2\alpha}^2} \sqrt{\hat{p}(1-\hat{p}) + \frac{z_{1/2\alpha}^2}{4n}} \right). \tag{8}$$

The Agresti–Coull Interval

If we start from (8), use its mid-point $\tilde{p} = \frac{n\hat{p} + 1/2z_{1/2\alpha}^2}{n + z_{1/2\alpha}^2}$ as a modified rate, and calculate the interval analogously to (6), we then obtain the *Agresti–Coull confidence interval*

$$\left(\tilde{p} - z_{1/2\alpha} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + z_{1/2\alpha}^2}}, \tilde{p} + z_{1/2\alpha} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + z_{1/2\alpha}^2}} \right). \tag{9}$$

This is related to the ‘adjusted Wald’ CI, defined according to the following advice for a 95% CI: ‘Add two successes and two failures and then use the Wald formula’.

Furthermore, their conclusion was that the Agresti–Coull interval dominates the other intervals in coverage but is, on average, slightly longer. The Wilson (1927) interval is comparable with the Agresti–Coull interval in both coverage and length. In addition, Brown et al. (2001) stressed that if one takes simplicity of presentation and ease of computation into account, the Agresti–Coull interval, although a bit too long, could be recommended for use. In our opinion, the Wilson CI is easy to interpret and is comparable with the Agresti–Coull CI in simplicity; we therefore recommend both as improved alternatives to the standard CI.

In studies of multiple maternities, the choice of method is important, for the conditions are conflicting. In general, one can expect a large number of maternities (n), but the proportion of multiple maternities (p) is small, and consequently, the distribution is skewed. In addition, in this study of the TWRs in small isolates the number of maternities can be moderate and sometimes even small, and consequently, the choice of the CI formula is of great importance.

Results

Our studies of the temporal trends in the countries of Norway, Sweden, and Finland show similar trends (Eriksson et al., 1995; Fellman & Eriksson, 1993; 2003; Fellman, 2016b). Comparisons are presented in Figure 1. One can observe similar levels and tendencies.

In our study to identify isolates with extreme TWRs, we constructed 99.9% CIs and compare them with the general TWRs. The chosen CIs are the Agresti–Coull ones. The high confidence levels were selected in order to stress the strong deviations and reduce the possibility that the differences are only random.

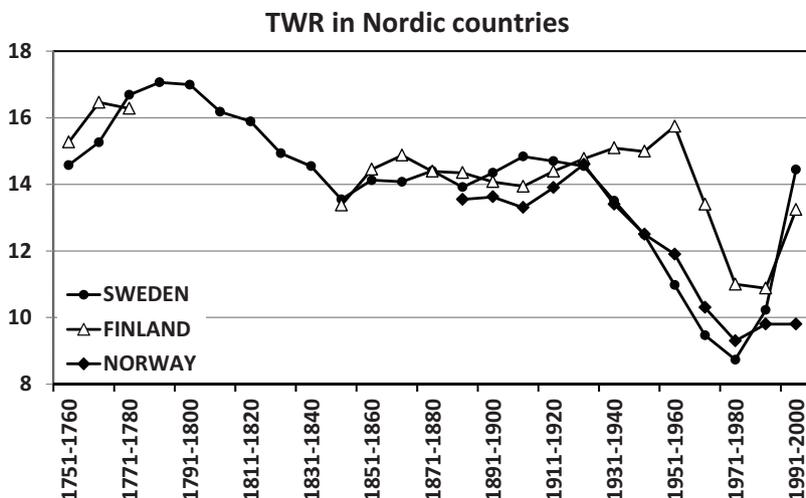


FIGURE 1
Comparisons of TWRs in Nordic countries. The temporal trends are similar and show marked fluctuations.

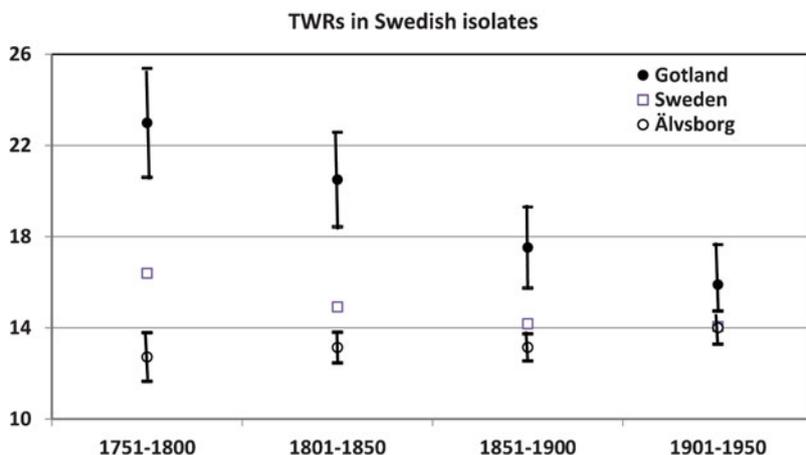


FIGURE 2
Comparison between the TWRs in Gotland, in the county of Älvsborg, and in Sweden overall. The discrepancies between the TWRs in the isolates and in Sweden are marked. The convergence between the TWRs is obvious.

Norway

Bonnevie and Sverdrup (1926) presented a detailed study of the TWRs in Norway. They paid special attention to extreme TWRs in an isolated Norwegian valley consisting of the three villages of Meldal, Rennebo, and Ringebu. They observed families with recurrent twin maternities. The family dataset consisted of 330 twin maternities among 10,155 maternities. The observed TWR was 32.50 per 1,000 and the 99.9% Agresti–Coull CI was (27.16, 38.82), indicating an extremely high TWR compared with the all-round TWR value of 14.67 given by Bonnevie and Sverdrup (1926) for Norway (1916–1917). They connected the observed high TWR in the valley to a strong heredity of twinning within the families.

Sweden

Earlier studies have found strong regional differences in the TWR. Extremely high TWRs have been registered for the island (county) of Gotland. Low TWRs have been identified in the county of Älvsborg located in the south-western part of Sweden (Fellman & Eriksson, 2003, 2009). In this study, we intended to identify statistically significant deviations in the local TWRs compared with the TWR for Sweden as a whole. We constructed CIs for the regional TWRs and compared them with the TWR for Sweden overall. In Figure 2, we present the results. We observed that in general neither the CIs for Gotland nor for Älvsborg cover the TWR for Sweden. This result indicates that the high TWRs for Gotland and the low TWRs for Älvsborg differ

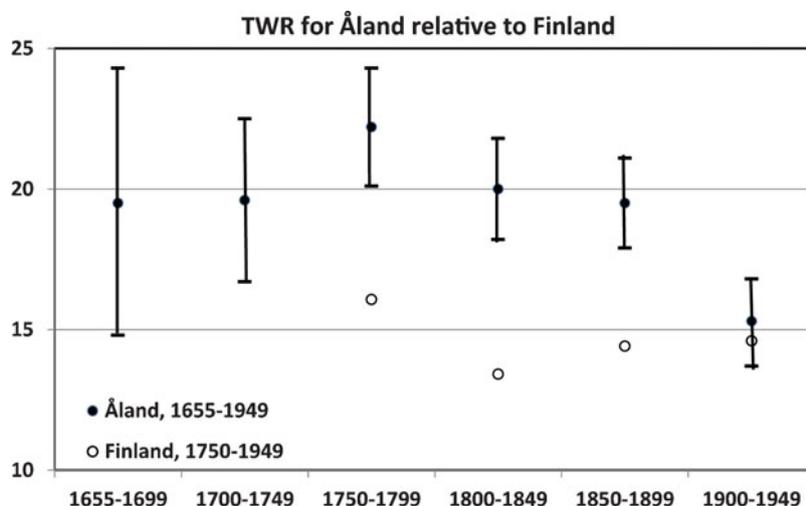


FIGURE 3

Comparison between TWRs for the Åland Islands and for Finland. The TWRs on the Åland Islands is markedly higher than those in Finland.

significantly from the general TWR for Sweden. Furthermore, one observes in Figure 2 that the TWR for both Gotland and Ålvsborg converge toward the TWR for Sweden. This last finding supports the convergence of the regional TWRs in Sweden found by Fellman and Eriksson (2005b).

Finland

In his classic study, Eriksson (1973) analyzed in detail the TWR for the Åland Islands. He especially stressed the high level of the TWR. In this study, we compared this high TWR and constructed CIs for the regional TWRs on the Åland Islands and compared them with the TWR for Finland overall. In Figure 3, we present the results. We observed that at least in the 20th century that the CIs for Åland do not cover the observed TWR for Finland. This indicates that the TWR for the isolated Åland Islands is significantly higher than the TWR for Finland.

Discussion

Strong geographical variations have been observed in the TWR. The TWR is high among people of African origin, intermediate among Europeans, and low among most Asiatic populations (Eriksson, 1973). In Europe, there tends to be a progressive increase in the TWR from south to north, with a minimum around the Basque provinces on the border between Spain and France. The highest TWRs in Europe have been noted among the Nordic populations (Bulmer, 1970; Eriksson, 1964, 1973; James, 1985).

Fellman and Eriksson (1990) examined the regional variation in the TWR in Finland for 1974–1983. Eriksson et al. (1995) presented a detailed study of the secular changes in the Nordic countries of Denmark, Finland, and Sweden. In our studies of the regional variation of the TWR in Sweden, we have analyzed TWRs for the different counties

(Eriksson & Fellman, 2004; Fellman & Eriksson, 2003, 2004, 2005a, 2009). In Fellman (2016a), the temporal variation in the Norwegian TWR was compared with corresponding trends in the neighboring Nordic countries of Iceland and Denmark.

Within larger populations, some small, isolated subpopulations have been identified to have extreme, mainly high, TWRs. Comparisons between the TWRs in isolates and the rates in the surrounding majority populations have been presented. These studies have recognized that the TWR often shows extreme values (high or low rates) within isolated subpopulations. Such investigations are often difficult to perform, partly because detailed local TWR data are not obtainable and partly because isolation details are not identifiable. This study had to be based on earlier investigations. Therefore, we considered the following isolates: the Åland Islands (Finland), the island of Gotland and the county of Ålvsborg (Sweden), and mountain villages (Norway).

Acknowledgments

This work was supported by grants from the Finnish Society of Sciences and Letters and the Magnus Ehrnrooth Foundation.

References

- Agresti, A., & Coull, B. A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *American Statistician*, 52, 119–126.
- Bonnevie, K., & Sverdrup, A. (1926). Hereditary predispositions to dizygotic twin-births in Norwegian peasant families. *Journal of Genetics*, 16, 125–188.
- Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statistical Science*, 16, 101–133.

- Brown, L. D., Cai, T. T., & DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Annals of Statistics*, 30, 160–201.
- Bulmer, M. G. (1970). *The biology of the twinning in man*. Oxford: Oxford University Press.
- Cramér, H. (1951). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.
- Eriksson, A. W. (1964). Pituitary gonadotrophin and dizygotic twinning. *Lancet* 2, 1298–1299.
- Eriksson, A. W. (1973). Human twinning in and around the Åland Islands. *Commentationes Biologicae*, 64, 1–159.
- Eriksson, A. W., Abbott, C., Kostense, P. J., & Fellman, J. O. (1995). Secular changes of twinning rates in Nordic populations. *Acta Geneticae Medicae et Gemellologiae (Roma)*, 44, 141–162.
- Eriksson, A. W., & Fellman, J. (2004). Demographic analysis of the temporal and regional variation in the rates of multiple maternities in Sweden since 1751. *Human Biology*, 76, 343–359.
- Fellman, J. (2013). Statistical analyses of mono- and dizygotic twinning rates. *Twin Research and Human Genetics*, 16, 1107–1111.
- Fellman, J. (2016a). Historic demography of Iceland. *British Journal of Medicine & Medical Research*, 2, 1–13.
- Fellman, J. (2016b). Temporal and spatial variations in the twinning rate in Norway. *Twin Research and Human Genetics*, 19, 359–366.
- Fellman, J. O., & Eriksson, A. W. (1990). Standardization of the twinning rate. *Human Biology*, 62, 803–816.
- Fellman, J. O., & Eriksson, A. W. (1993). Biometric analysis of the multiple maternities in Finland, 1881–1990 and in Sweden since 1751. *Human Biology*, 65, 463–479.
- Fellman, J., & Eriksson, A. W. (2003). Temporal differences in the regional twinning rates in Sweden after 1750. *Twin Research*, 6, 183–191.
- Fellman, J., & Eriksson, A. W. (2004). Should the statistical analyses of twinning-rate data be improved? *Twin Research*, 7, 398–405.
- Fellman, J., & Eriksson, A. W. (2005a). Variations in the maternal-age effect on twinning rates: The Nordic experience. *Twin Research and Human Genetics*, 8, 515–523.
- Fellman, J., & Eriksson, A. W. (2005b). The convergence of the regional twinning rates in Sweden, 1751–1960. *Twin Research and Human Genetics*, 8, 163–172.
- Fellman, J., & Eriksson, A. W. (2006). Weinberg's differential rule reconsidered. *Human Biology*, 78, 293–315.
- Fellman, J., & Eriksson, A. W. (2009). Spatial variation in the twinning rate in Sweden, 1751–1850. *Twin Research and Human Genetics*, 12, 583–590.
- James, W. H. (1985). Dizygotic twinning, birth weight and latitude. *Annals of Human Biology*, 12, 441–447.
- Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of American Statistical Association*, 22, 209–212.