# Should the Statistical Analyses of Twinning-Rate Data be Improved?

Johan Fellman and Aldur W. Eriksson

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

Every statistical model is based on explicitly or implicitly formulated assumptions. In this study we address new techniques of calculation of variances and confidence intervals, analyse some statistical methods applied to modelling twinning rates, and investigate whether the improvements give more reliable results. For an observed relative frequency, the commonly used variance formula holds exactly with the assumptions that the repetitions are independent and that the probability of success is constant. The probability of a twin maternity depends not only on genetic predisposition, but also on several demographic factors, particularly ethnicity, maternal age and parity. Therefore, the assumption of constancy is questionable. The effect of grouping on the analysis of regression models for twinning rates is also considered. Our results indicate that grouping influences the efficiency of the estimates but not the estimates themselves. Recently, confidence intervals for proportions of low-incidence events have been a target for revived interest and we present the new alternatives. These confidence intervals are slightly wider and their midpoints do not coincide with the maximum-likelihood estimate of the twinning rate, but their actual coverage is closer to the nominal one than the coverage of the traditional confidence interval. In general, our findings indicate that the traditional methods are mainly satisfactorily robust and give reliable results. However, we propose that new formulae for the confidence intervals should be used. Our results are applied to twin-maternity data from Finland and Denmark.

Every statistical analysis is based on explicitly or implicitly formulated assumptions that are intended to form a stochastic model of the problem. The model should be as simple as possible, while still giving a realistic description and interpretation of the problem. If the model is too simple, the goodness of fit may not be satisfactory. A more complicated model gives a better fit, but the analyses and interpretations may be difficult. In this study, we analyse some statistical methods and investigate whether some improvements give more reliable results.

For an observed relative frequency, the variance formula commonly used holds exactly only under the assumptions (a) that the repetitions are independent and (b) that the probability of success is constant. Usually there are no discussions about these premises. Cramér (1951, p. 206) studied the effect of defects in the assumption (b) on the accuracy of the variance formula. He also stated that Poisson had already considered the problem in the 19th century. The probability of a twin maternity depends on several demographic factors, particularly ethnicity, maternal age and parity (birth order), and also shows temporal and regional variations within countries. Even when these factors are considered, interindividual heterogeneity can be expected. Assumption (b) cannot hold exactly for twinning rates (TWRs) and, therefore, we checked the robustness of the variance formula. If we consider the data grouped in homogeneous classes, the standard variance formula can be split into two parts. The first measures the intragroup and the second the intergroup variation. Both our theoretical and numerical studies indicate that the standard formula is sufficiently exact and, for most cases, corrections can be disregarded. Furthermore, we note that if the heterogeneity is ignored, then the variance is overestimated.

A common situation, when the effect of influential factors such as maternal age and parity on the TWRs are analysed (by regression models), is that the observations are divided into groups in which the factors are constant. The effect of this grouping on the analysis of regression models for the TWRs is considered. Recently we have analysed temporal and regional variations in the TWR for Sweden, 1751–1960. We have tried to explain the variations with different demographic and socioeconomic factors (Fellman & Eriksson, 2003; Eriksson & Fellman, 2004).

*Received 29 March, 2004; accepted 3 June, 2004.*

*Address for correspondence: Professor Johan Fellman, Folkhälsan Institute of Genetics, Population Genetics Unit, POB 211, FIN–00251 Helsinki, Finland. E-mail: johan.fellman@shh.fi*



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

Recently, confidence intervals (CIs) for the proportions of low-incidence events have been a target for revived interest (e.g., Brown et al., 2001; 2002)). The alternative CIs obtained are slightly wider and their midpoints do not coincide with the maximum-likelihood (ML) estimate of the rate. However, the general theory confirms that the actual coverage for the alternative CI is closer to the nominal one than to the coverage of the CI based on the standard formula. We present these new findings and use them when we compare variations in the TWRs.

## Methods

### Variance of Observed Proportions

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

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

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, 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 maternal age, parity and ethnicity. Hence, the variance formula (1) for the total twinning rate 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 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 let 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 $r$. 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

$$\hat{p} = \frac{1}{n}\sum_r n_r \hat{p}_r = \frac{1}{n}\sum_r S_r = \frac{S}{n} \quad (2)$$

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)

$$Var(\hat{p}) = \frac{1}{n^2}\sum_r n_r^2 Var(\hat{p}_r) = \frac{1}{n^2}\sum_r n_r^2 \frac{p_r(1-p_r)}{n_r} = \frac{p(1-p)}{n} - \frac{1}{n^2}\sum_r n_r (p_r - p)^2 \,. \quad (3)$$

This fundamental result can also be written

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

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 $Var\ \hat{p}$ 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 intragroup 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 intergroup variation and also in this case the formula (1) holds.

The results obtained have an interesting interpretation. Without any grouping, the variance formula (1) holds. A grouping giving additional information



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

about the variation of the probability reduces the variance. The reduction from (1) to

$$Var(\hat{p}) = \frac{1}{n^2}\sum_r n_r p_r (1 - p_r)$$

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. According to the examples presented below, the grouping effect is, as a rule, minute (cf. Table 1).

The analysis performed above indicates that both the classical variance formula and the variance formula (3) are overestimates 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 interindividual differences in the probabilities. This overestimation indicates that all statistical tests give too-low statistical values and the tests are conservative.

#### The Effect of Grouping on the Estimation of Models for the TWR

When models are built for TWRs, the common situation is that the observations fall into groups, where the influential factors are constant. Let the number of groups be $R$ and the number of maternities be $n_r$ ($r = 1,\ldots,R$). In an earlier paper (Fellman & Eriksson, 1987), we discussed weighted least squares (WLS) estimation in connection with twin studies. Following our notations in this article, we suggested in that paper that the weights would be $n_r$ ($r = 1,\ldots,R$). We also discussed the influence of the true TWRs on the weights, but there, as well as here, we stressed the difficulties of including the unknown rates in the weights. In addition, we stressed that ignoring the rates has little influence on the efficiency of the estimates. Maddala (1985) has suggested a three-stage estimation. First, an ordinary least square (OLS) estimation, then an estimation of the variances and after that a weighted least square (WLS) estimation with the estimated variances as weights. In our opinion, a twin maternity is a relatively rare phenomenon and the variance shows such small variations that it is not necessary to apply Maddala's attempt in order to improve the efficiency. In addition, the constancy obtained below of the parameter estimates for the two models speaks in favor of the simpler method.

In this article we consider the outcome $y_i$ of the maternity number $i$, and define

$$y_i = \begin{cases} 1 & \text{if the outcome is a twin set} \\ 0 & \text{else} \end{cases} \qquad (5)$$

We start the analysis with the model defined for the individual outcomes

$$y_i = \alpha + \beta_1 x_{i1} + \ldots + \beta_m x_{im} + \varepsilon_i \qquad i = 1,\ldots,n \qquad (6)$$

where the values of $y_i$, at this stage, may be more generally defined than in (5) and $x_{ij}$ ($i = 1,2,\ldots,n$, $j = 1,\ldots m$) are the values of the influential factors (maternal age, parity, year of birth, etc.) for mother number $i$. We assume that the error terms $\varepsilon_i$ are independent and have the properties $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. In fact, $Var(\varepsilon_i) = p_i(1 - p_i)$, where $p_i$ is the probability for a twin set for the maternity number $i$. Hence, $Var(\varepsilon_i)$ varies but is fairly constant, and the loss in efficiency can be considered minute.

Now we assume that our observations fall into groups where the regressors are constant. Consequently, we obtain the model

$$\bar{y}_r = \alpha + \beta_1 x_{r1} + \ldots + \beta_m x_{rm} + \bar{\varepsilon}_r \qquad r = 1,\ldots,R \qquad (7)$$

where

$$\bar{y}_r = \frac{1}{n_r}\sum_{i=1}^{n_r} y_{ir}\,,\ E(\bar{\varepsilon}_r) = 0\,,\ Var(\bar{\varepsilon}_r) = \frac{1}{n_r}\sigma^2 \text{ and } x_{r1},\ldots,x_{rm}$$

are the (mean) values of the regressors in group number $r$.

If the observations in (7) are weighted with the weights $n_r$ ($r = 1,\ldots,R$), model (6) and (7) give the same parameter estimates. The reduction in the total sum of squares according to the grouping is the intragroup variation and, after the grouping, the variation within the groups is ignored. When the proposed methods are applied to the TWR data, we obtain the following results. Let $N_r$ and $n_r$ be the number of maternities and twin maternities in group number $r$, respectively. Let

$$\bar{p}_r = \frac{n_r}{N_r} \text{ and } \hat{p}_r$$

be the observed and expected TWRs in group number r, respectively. Let $N$, $n$ and

$$\bar{p} = \frac{n}{N}$$

be the observed number of maternities, twin maternities and the total TWR for the whole data set.

The mean sum of squares is

$$MS_{RU} = s_U^2 = \frac{1}{N-k-1}\sum_{r=1}^{R}\sum_{i=1}^{n_r}(p_{ir} - \hat{p}_r)^2 \qquad (8)$$

which is necessary for the estimation of the standard errors (SEs) of the parameter estimates. Based on the grouped data, the corresponding mean sum of squares is

$$MS_{RG} = s_G^2 = \frac{1}{R-k-1}\sum_{r=1}^{R} N_r(\bar{p}_r - \hat{p}_r)^2 \qquad (9)$$

A comparison of formulae (8) and (9) gives the relative efficiency between the estimations based on ungrouped and grouped data. The efficiency of the parameter estimates depends on the reduction in the variation compared to the reduction in the degrees of freedom, that is $N - R$. If the reduction in the variance is small compared to the reduction in the degrees



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

of freedom, the residual sum of squares for the grouped data is large in comparison with the residual sum of squares for the ungrouped data and the ungrouped data give more efficient estimates. The opposite situation is also possible.

A regression analysis based on ungrouped data is, as such, often difficult to perform. This is a consequence of the extensive data set. For twinning data, the following method can be used, because the regressand can only obtain the values one (a twin maternity) or zero (not a twin maternity) and the ungrouped data set can be reconstructed from the grouped data. We estimate the parameters using the grouped data and obtain the common parameter estimates. The choice between ungrouped and grouped data sets should be based on the corresponding efficiency of the parameter estimate. However, this method is not possible in the general case, because in general one cannot distinguish the individual observations $y_{ir}$ from the general group mean $\bar{y}_r$. Upon request, the interested reader can obtain from the authors a detailed mathematical presentation of the results in this section.

#### Confidence Interval (CI) for an Unknown Proportion

Now, we consider the CIs for an unknown proportion $p$. The ML estimator of $p$ is

$$\hat{p} = \frac{S}{n}$$

where $S$ is the number of twin maternities and $n$ is the total number of maternities. It is a well-known fact that $\hat{p}$ has the mean $p$ and is asymptotically normal. If we consider ungrouped data or data for a specific group, the variance is

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

The standard method, at least in applied studies, is that the variance formula is estimated by replacing $p$ with $\hat{p}$, resulting in the approximate test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \tag{10}$$

Based on (10), the $100(1-\alpha)$% CI, the so-called Wald's confidence interval, is

$$\left(\hat{p} - z_{½\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{½\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \tag{11}$$

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

Recently, Brown et al. (2001; 2002) gave a thorough presentation of the problems concerning CIs for unknown proportions. They showed that if the standard method (11) is applied, the actual coverage probability for $p$ can differ markedly from the nominal confidence level at realistic and even larger than realistic sample sizes. The error arises from two sources: discreteness and skewness ($p \neq .5$) in the underlying distribution. For a two-sided interval, the rounding error due to discreteness is dominant, being of order $n^{-1/2}$. The error due to skewness is of the order $n^{-1}$, but is still important for even a moderately large $n$. Although

$$\hat{p} = \frac{S}{n}$$

is an unbiased ML estimator of $p$, as the centre of a CI it causes a systematic negative bias in the coverage. Especially, Brown et al. (2001) have stressed that, even in cases when the textbooks indicate that the formula (11) is safe, the coverage probability may differ markedly from that expected. As alternatives they proposed the Agresti-Coull (1998) or the Wilson (1927) interval. The alternative methods are briefly presented below.

***The Wilson interval.*** Using the notations introduced above, we consider the exact test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \tag{12}$$

which, on theoretical grounds, can be considered more reliable than the approximate one in (10). From (12) we obtain after some calculations the $100(1-\alpha)$ % Wilson CI

$$\left(\frac{n\hat{p} + ½z^2_{½\alpha}}{n + z^2_{½\alpha}} - \frac{z_{½\alpha}\sqrt{n}}{n + z^2_{½\alpha}}\sqrt{\hat{p}(1-\hat{p}) + \frac{z^2_{½\alpha}}{4n}},\ \frac{n\hat{p} + ½z^2_{½\alpha}}{n + z^2_{½\alpha}} + \frac{z_{½\alpha}\sqrt{n}}{n + z^2_{½\alpha}}\sqrt{\hat{p}(1-\hat{p}) + \frac{z^2_{½\alpha}}{4n}}\right) \tag{13}$$

***The Agresti-Coull interval.*** If we start from (13), use its midpoint

$$\tilde{p} = \frac{n\hat{p} + ½z^2_{½\alpha}}{n + z^2_{½\alpha}}$$

as a modified rate, and calculate the interval analogously to (11), then we obtain the Agresti-Coull interval

$$\left(\tilde{p} - z_{½\alpha}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + z^2_{½\alpha}}},\ \tilde{p} + z_{½\alpha}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + z^2_{½\alpha}}}\right) \tag{14}$$

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, Brown et al. (2001, 2002) concluded that the Agresti-Coull (1998) interval dominates the other intervals in coverage, but is on average slightly longer. The Wilson (1927) interval is comparable with the Agresti-Coull (1998) 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 interval is easy to interpret and is comparable with the Agresti-Coull interval in simplicity and therefore we recommend these two



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

as improved alternatives to the standard CI. In studies of multiple maternities, the choice of method is important, for the conditions are conflicting. One can expect a large number of total maternities ($n$), but the proportion of multiple maternities ($p$) is small.

## Results

### The Reliability of the Variance Formula (1)

Our decision concerning the robustness of the variance formula (1) is based on some empirical examples. The first example is Danish data (1931–1967) grouped according to year of birth, maternal age, parity and marital status. The Danish data for the period 1931–1967 consist of 2,854,887 maternities, including 38,661 twin sets, yielding the total TWR 13.54 per 1000. The second example is based on twinning data from Finland, 1953–1964, grouped according to maternal age, parity and marital status (Eriksson & Fellman, 1967a, 1967b). The Finnish data consist of 1,019,619 maternities, including 15,467 twin sets, yielding the total TWR 15.17 per 1000. The statistical results are given in Table 1 and we observe that the correction components are rather small and can in general be ignored. For the Finnish data we present in addition the effects of the individual grouping factors. As a consequence of the correlations between the factors, the individual effects do not add to the joint effect of all factors. The intergroup variation is relatively larger in the Danish data than in the Finnish data. For the Danish data the grouping has small but discernible effects, but for the Finnish data the effect can be disregarded. According to the theoretical analyses above this difference indicates that the factors considered are more influential for the Danish than for the Finnish data.

In these examples we observe that the corrections are minute, but in studies where one considers small series, great differences between the 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. Our findings indicate that the corrections recommended from a theoretical point of view are small and can be disregarded in empirical statistical analyses.

### Model Building of TWR

***Finnish data, 1953–1964.*** The proposed analyses are applied to the Finnish data for the period 1953–1964 presented in Eriksson and Fellman (1967a; 1967b) and analysed in the previous section. However, now we only consider the age groups less than 40 years, because for such data the TWR is (approximately) linearly dependent on maternal age and parity and we have the possibilities to build a linear regression model where the regressors are maternal age (AGE), parity (PARITY) and marital status (MARITAL: married mothers = 0, unmarried mothers = 1) (Fellman & Eriksson, 1987). The regression model is

$$TWR = \alpha + \beta_1 AGE + \beta_2 PARITY + \beta_3 MARITAL + \varepsilon$$

This model is applied to both the individual data and the grouped data. The parameter estimates and their SEs are given in Table 2. All the parameter estimates are statistically significant. It is a well-known fact that maternal age and parity have significant influences on the TWR, but the effect of marital status varies in different studies. In this case, the reduction in the sum of squares due to the grouping is less than the mean residual sum of squares for the grouped data. Consequently, the ungrouped data set gives more efficient estimates.

**Table 1**

The Effect of the Grouping on the Efficiency of the Estimated TWR

The components of the variation are given in per cent. The Danish data for the period 1931–1967 are grouped according to the year of birth, maternal age, parity and marital status. The Finnish data for the period 1953–1964 are grouped according to maternal age, parity and marital status. For Finland the individual effects of these factors are also presented. As a consequence of the correlations between the factors, the individual effects do not add to the joint effect of all factors. The results are presented in more detail in the text.

| Population and grouping factors | Intragroup variation $\frac{1}{n^2}\sum_r n_r p_r(1-p_r)$ | Intergroup variation $\frac{1}{n^2}\sum_r n_r (p_r - p)^2$ | Total variation $\frac{p(1-p)}{n}$ |
|---|---|---|---|
| Denmark, 1931–1967 | | | |
| All factors | 99.17 | 0.83 | 100.000 |
| Finland, 1953–1964 | | | |
| Maternal age | 99.84 | 0.16 | 100.000 |
| parity | 99.91 | 0.09 | 100.000 |
| marital status | 99.99995 | 0.00005 | 100.00000 |
| All factors | 99.81 | 0.19 | 100.000 |



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

**Table 2**

Parameter Estimates Obtained for Ungrouped and Grouped Data Sets in Finland 1954–1963

The regression model is built for TWR (defined per 1000). We observe that the estimates are identical, but the SEs differ, indicating that the ungrouped data set gives more efficient estimates. All the parameter estimates are statistically significant and consequently, the regression model fits the data well.

| | Ungrouped | | Grouped | |
|---|---|---|---|---|
| | Parameter estimation | | | |
| Parameter | Estimate | *SE* | Estimate | *SE* |
| Intercept | –6.939 | 0.666 | –6.939 | 0.888 |
| Age | 0.744 | 0.027 | 0.744 | 0.036 |
| Parity | 0.895 | 0.086 | 0.895 | 0.114 |
| Marital | 3.351 | 0.622 | 3.351 | 0.830 |

***Danish data, 1931–1967.*** We also consider the Danish data for the period 1931–1967 (Fellman & Eriksson, 2002) and analyzed in the previous section. Again we consider only age groups less than 40 years. We build a linear regression model as for the Finnish data, but also including the year of birth (YEAR). Consequently, we consider the regression model

$$TWR = \alpha + \beta_1 AGE + \beta_2 PARITY + \beta_3 MARITAL + \beta_4 YEAR + \varepsilon$$

The parameter estimates and their *SE*s are given in Table 3. The parameter estimates for year of birth, maternal age and parity, but not for marital status, are statistically significant. The year of birth parameter estimate indicates a significant time trend. We observe that the estimates for ungrouped and grouped data sets are identical, but the *SE*s differ. In this case the reduction in the sum of squares due to the grouping is less than the mean residual sum of squares for the grouped data. Consequently, the ungrouped data set yields more efficient estimates.

For the Finnish data, there is a statistically significant effect of marital status. This is in agreement with our findings in Eriksson and Fellman (1967a; 1967b). In the Danish data, however, the marital status effect is not statistically significant, which we have already observed in Fellman and Eriksson (1987, 2002). The explanation for this is that the effect of marital status on the TWR is diminishing. We are still of the opinion that this trend is caused by the fact that the classification of marital status started to lose its initial meaning after the middle of the 20th century with the resurgence of feminism in the 1960s, and a greater acceptance of premarital sex. Today in Denmark particularly younger couples very often live as married couples with no formal wedding. Consequently, they are statistically classified as unmarried. During the period 1973–1984 the proportion of maternities among unmarried mothers in Denmark increased from 17.2 to 42.1 % (Fellman & Eriksson, 1987).

Finally, we want to stress that the ungrouped data sets are very large and therefore difficult to analyze in a straightforward way, and the grouped data are to be preferred. However, if the ungrouped data give more efficient estimates within this special framework, analyses based on ungrouped data can be performed, as indicated earlier.

### Temporal Trends in the TWR for Åland, 1653–1960

In the following example we apply the new CIs and compare them with the standard CI. Eriksson (1973) presented the TWR for Åland for the period 1653–1959. In the 17th century, the number of maternities per decade was on average about 650 and, in the 20th century, the corresponding number

**Table 3**

Parameter Estimates Obtained for Ungrouped and Grouped Data Sets in Denmark, 1931–1967

The regression model is built for TWR (defined per 1000). The parameter estimates for year of birth, maternal age and parity but not for marital status are statistically significant. We observe that the estimates are identical but the SEs differ, indicating that the ungrouped data set gives more efficient estimates.

| | Ungrouped | | Grouped | |
|---|---|---|---|---|
| | Parameter estimation | | | |
| Parameter | Estimate | *SE* | Estimate | *SE* |
| Intercept | 0.770 | 0.541 | 0.770 | 0.616 |
| Year | 0.076 | 0.007 | –0.076 | 0.008 |
| Age | 0.568 | 0.016 | 0.568 | 0.018 |
| Parity | 0.909 | 0.052 | 0.909 | 0.059 |
| Marital | 0.424 | 0.255 | 0.424 | 0.290 |



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

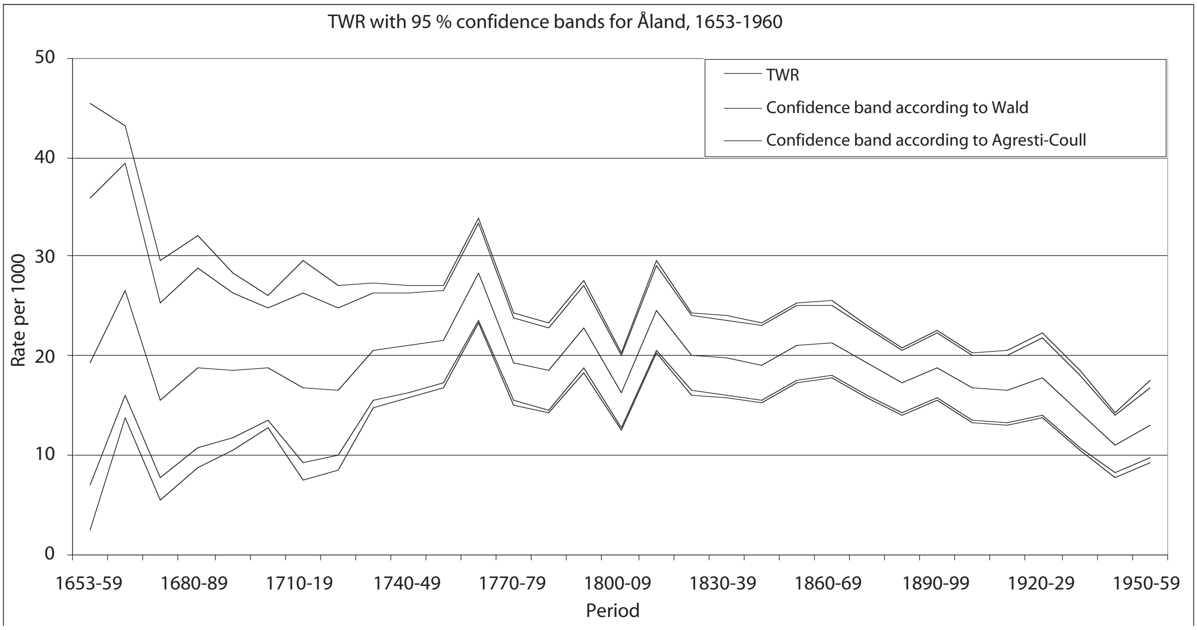


**Figure 1**

The TWR for Åland for the period 1653–1959, according to Eriksson (1973).
In the 17th century the number of maternities per decade was on average about 650, and in the 20th century the corresponding number was about 4380. In the figure we have included 95% confidence bands based on the Wald and the Agresti-Coull confidence intervals. The effect of the sample sizes on the differences in the CIs is apparent.

was about 4380. In Figure 1 we present the temporal trend in the TWR and the 95% confidence bands based on the Wald and the Agresti-Coull formulae (11) and (13), respectively. The Wilson interval is not included in the figure, because the difference between the Agresti-Coull interval and the Wilson interval should not be discernible. The effect of the sample sizes on the differences in the CIs is apparent.

## Discussion

If we reconsider the results in the examples we noted that for the Finnish data there is a statistically significant effect of marital status. This is in agreement with our findings in Eriksson and Fellman (1967a; 1967b). In the Danish data, however, the marital status effect is not statistically significant, which we have already observed in Fellman and Eriksson (1987; 2002). The explanation for this is that the effect of marital status on the TWR is diminishing. We are still of the opinion that this trend is caused by the fact that the classification of marital status started to lose its initial meaning after the middle of the 20th century. Today younger couples very often live as married couples, but they are statistically classified as unmarried.

Summing up, we can conclude that the analyses of the statistical methods considered, indicate that the corrections recommended from a theoretical point of view, for the variance formula are small and can be disregarded in empirical statistical analyses. Based on the comparison between model-building based on ungrouped and grouped data sets we have found strong evidence that there are no marked advantages for one of them. This finding supports the established model-building methods found in the literature. The choice between ungrouped and grouped data sets should be based entirely on the corresponding efficiency of the parameter estimate. The only method which we stress should be changed is the use of confidence intervals. Now, we strongly recommend the introduction of the new CIs presented and analysed above.

## Acknowledgments

This work was supported by grants from the 'Liv och Hälsa' Foundation and from The Finnish Society of Sciences and Letters. We are very grateful to the Danish Statistical Bureau for cooperation in sending us information on details of twin maternities.

## References

Agresti, A., & Coull, B. A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *American Statistician, 52,* 119–126.

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.



https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press

Cramér, H. (1951). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

Eriksson, A. W. (1973). Human twinning in and around the Åland Islands. *Commentationes Biologicae, 64,* 1–154.

Eriksson, A. W., & Fellman, J. (1967a). Twinning and legitimacy. *Hereditas, 57,* 395–402.

Eriksson, A. W., & Fellman, J. (1967b). Twinning in relation to the marital status of the mother. *Acta Genetica, 17,* 385–398.

Eriksson, A. W., & Fellman, J. (2004). Demographic analysis of the temporal and regional variation in the rates of multiple maternities in Sweden since 1749. *Human Biology* (in press).

Fellman, J. O., & Eriksson, A. W. (1987). Statistical models for the twinning rate. *Acta Geneticae Medicae et Gemellologiae, 36,* 297–312.

Fellman, J., & Eriksson, A. W. (2002). Standardisation of the twinning rate. *Twin Research, 5,* 19–29.

Fellman, J., & Eriksson, A. W. (2003). Temporal differences in the regional twinning rates in Sweden after 1750. *Twin Research, 6,* 183–191.

Maddala, G. S. (1985). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of American Statistical Association, 22,* 209–212.




https://doi.org/10.1375/twin.7.5.398 Published online by Cambridge University Press