

# Number of lethal equivalents in human populations: how good are the previous estimates?

JAE KYUN LEE\*, MARTIN LASCOUX† & ERIK V. NORDHEIM‡

*Department of Mathematics, University of Southern California, DRB 273, 1042 W. 36th Place, Los Angeles, CA 90089, U.S.A., †Department of Plant Breeding, Swedish University of Agricultural Sciences, Uppsala Genetic Centre, Box 7003 750 07 Uppsala, Sweden and ‡Departments of Statistics and Forestry, University of Wisconsin–Madison, 1630 Linden Dr, Madison, WI 53706, U.S.A.*

Lee *et al.* (1996) recently developed a method for interval estimation of the number of lethal equivalents by using a hierarchical structure of likelihood functions. This hierarchical model consists of two multinomial trials: one of the sampling process of the parents from the population of interest, and the other for the survival of the offspring of the families obtained by mating the parents. The method, initially developed for selfing and full-sib mating, is extended here to include more general mating systems as well as mixtures of mating systems. We applied it to human data sets for which confidence intervals were previously not available. Our point estimates were close to previous ones, and the standard deviations were generally quite small. Thus, even if debate over the meaning of the concept of lethal equivalents has not been entirely resolved, our results showed that the previous estimates are at least statistically meaningful.

**Keywords:** consanguineous marriages, deleterious alleles, inbreeding coefficient, lethal equivalents, marginalized methods of moments.

## Introduction

In 1956, Morton *et al.* defined the concept of lethal equivalents, proposed a method to estimate it and applied it to three human data sets then available. The influence of this paper, called ‘the classic paper of the theory of genetic loads’ by Cavalli-Sforza & Bodmer (1971), on subsequent studies of deleterious genes has been quite substantial. Although first used on human populations (Morton *et al.*, 1956; Morton, 1975; Rao & Inbaraj, 1977; Vogel & Motulsky, 1982; Khat, 1988), the proposed estimation method has now been applied to a wide range of organisms, including pine trees (Sorensen, 1969), willow trees (Kang *et al.*, 1994), ferns (Klekowski, 1988), bees (Kerr, 1975), and endangered animal species (Ralls *et al.*, 1988). From their study, Morton *et al.* concluded that ‘the average person carries about four to five genes which, if homozygous, could cause conspicuous abnormality’. Later studies often led to lower values, although values of the estimates varied

greatly among populations. This variation was used by some authors to question the validity of the concept of lethal equivalents as well as the assumptions used in the model (Vogel & Motulsky, 1982). However, even if the conditions of estimation had been uniform, it was difficult to interpret the estimates in the absence of a rigorous measure of variability. A maximum likelihood estimate of the number of lethal equivalents can be obtained by a weighted regression of mortality on inbreeding level,  $F$  (for example, Morton *et al.*, 1956; Morton, 1960; Chakraborty & Chakravarti, 1977; Ralls *et al.*, 1988). This estimate, however, suffers from at least two limitations: first, it is derived under the assumptions required for regression models which may not be relevant if our goal is to estimate the number of lethal equivalents; secondly, when the number of inbreeding levels is small and limited to a narrow range of  $F$  values, which is often the case, the estimate is not reliable. Therefore, a direct statistical estimation method would be most useful.

More than 40 years have now passed, and the debate on human genetic loads has abated; some

\*Correspondence.

discussion can be found only in the historical assessment of the last decades (Gillespie, 1991; Wallace, 1991; Crow, 1993; Ewens, 1993). We wish to return to the first estimates of Morton *et al.* (1956) and ask: how good are these estimates? Our purpose here is to do just that, based on the interval estimation of the number of lethal equivalents that is now available (Lee *et al.*, 1996). First, we recall the definition of lethal equivalents and describe the method of Lee *et al.* (1996) in detail for half-sib mating. Then we extend that method to other mating designs and various types of consanguineous marriages in humans. We also suggest an estimation method for a mixture of data sets from several different mating designs (or different types of consanguineous marriages). We illustrate our method with several data sets from previous studies.

### Estimation methods for the number of lethal equivalents

The number of lethal equivalents represents the magnitude of the overall effect of deleterious genes (Morton *et al.*, 1956; Cavalli-Sforza & Bodmer, 1971). In a diploid population, in the absence of epistasis, the number of lethal equivalents per gamete is:

$$\varepsilon = \sum_{k=1}^M s_k q_k, \quad (1)$$

where  $M$  is the number of loci with segregating deleterious alleles,  $q_k$  is the deleterious allele frequency at the  $k$ th locus and  $s_k$  is the selection coefficient. At the zygotic level, the number of lethal equivalents is  $2\varepsilon$ ; let us define  $E = 2\varepsilon = 2\sum s_k q_k$ . The selection coefficient  $s$  is defined from the convention that the probabilities of survival of dominant homozygote, heterozygote, and recessive homozygote are 1,  $1-hs$ , and  $1-s$ , respectively, where  $h$  is the coefficient of dominance.

The classical method (MCM for Morton–Crow–Muller) of estimating lethal equivalents was derived by defining the probability of eventual survival ( $S$ ) of an individual to be the product of the probabilities of the individual surviving different agents that could cause the mortality. This assumes that various causes of mortality, such as lethal genes at different loci and environmental effects, act independently. Based on these assumptions, MCM derives a linear relationship between  $\log S$  and the inbreeding coefficient ( $F$ ) so that

$$-\log S = A + BF, \quad (2)$$

where  $A = \sum x + \sum q^2 s + 2\sum q(1-q)sh$ ,  $B = \sum qs - \sum q^2 s - 2\sum q(1-q)sh$ , and  $x$  represents environmental causes of death. From calculating  $A$  and  $B$ , Morton *et al.* (1956) approximated the number of lethal equivalents ( $\varepsilon$ ), where  $\varepsilon$  is the first term of  $B$ . The MCM method assumes that the sampling is made from a large population with large  $M$  and small  $q$ . We note that if the original population is small or inbred, this assumption does not hold, and this method is likely to result in biased estimates because such a logarithmic approximation in eqn 2 cannot be assumed (Kang *et al.*, 1994). The MCM method does not appear to lend itself to direct statistical estimation of the number of lethal equivalents.

An alternative method (COMB for combinational method), which uses a combinatorial expression of the probability of mortality ( $Q$ ), has been widely used in the forest genetics literature for selfing (Bishir & Namkoong, 1987) and full-sib mating (Kang *et al.*, 1992). Contrary to MCM, which makes inferences about the entire population, the COMB method makes inferences about a particular individual; forest geneticists have used the average of the individual numbers of lethal equivalents to make inferences about the entire population.

Both MCM and COMB methods have been adapted to permit the removal of environmental causes of mortality (Bishir & Namkoong, 1987; Savolainen *et al.*, 1992). Makov & Bittles (1986) proposed an approach based on regression techniques to examine the error distribution of estimates of  $E$  obtained by modifying the model underlying the MCM estimation method. Besides the questionable genetic meaning of these alternative models, such variation does not represent the actual deviation of their estimate from the true value of  $E$ . Recently, Lee *et al.* (1996) developed a method for obtaining interval estimates of the number of lethal equivalents by using a hierarchical structure of likelihood functions. This hierarchical model consists of two multinomial trials: one for the sampling process of the parents from the population of interest and the other for the survival of the offspring of the families obtained by mating the parents. This model and its statistical aspects were described for two commonly used mating systems in plant genetics, selfing and full-sib, and environmental effects on mortality were incorporated. Because of unobservable individual genotypes in mating experiments and the complexity of the corresponding likelihood function, Lee *et al.* (1996) used a variant of methods of moments, *marginalized methods of moments*, summing out the unobserved genotype variables and then evaluating the moment expectations, for

obtaining their statistical estimates of the number of lethal equivalents. This method is based on a large number of independent families with the same inbreeding level and under the same environment, where the common environment effect is taken into account by assuming that the mortality in a noninbred control is entirely the result of environmental causes; this can be extended to a more general situation of nonconstant environments by estimating these effects from the noninbred control group of each subpopulation (unpublished result).

**Extension of the method to general single mating systems**

The same hierarchical model can be extended to mating systems other than the selfing and full-sib matings in Lee *et al.* (1996), including various consanguineous marriages observed in human populations. The half-sib case, when mates have only one common parent, will be used in the following to illustrate how this can be achieved. The sampling distribution of the parental generation can be defined by the number of all possible combinations of deleterious loci that can lead to a homozygous offspring for any of these loci. Hence, for any given locus, assuming that there are only two types of alleles at each locus (a deleterious allele *a* and a nondeleterious one *A*) the number of possible combinations of loci carrying deleterious gene(s) among the three parents is 26 (= 3<sup>3</sup> - 1, as there are three possible genotypes, *AA*, *Aa* and *aa*, for each of the three parents). However, among these 26 combinations, considering only two parental genotypes (*AA* and *Aa*), only seven (= 2<sup>3</sup> - 1) (*m*<sub>1</sub>, ..., *m*<sub>7</sub>) are relevant here. These seven combinations of loci correspond to five situations: (i) the three parents are heterozygous (*m*<sub>1</sub>), (ii) the common parent as well as one of the nonshared parents are heterozygous (*m*<sub>2</sub> and *m*<sub>3</sub>), (iii) both nonshared parents are heterozygous (*m*<sub>4</sub>), (iv) only the common parent is heterozygous (*m*<sub>5</sub>), and (v) only one of the nonshared parents is heterozygous (*m*<sub>6</sub> and *m*<sub>7</sub>). Sixteen other combinations can also lead to homozygous offspring, but are not included here because the frequency of homozygotes of deleterious alleles, *q*<sup>2</sup>, is likely to be extremely low in the population. The vector (*m*<sub>1</sub>, ..., *m*<sub>7</sub>) follows a multinomial distribution with parameters *M*,  $\xi^3$ ,  $\xi^2(1-\xi)$ ,  $\xi^2(1-\xi)$ ,  $\xi^2(1-\xi)$ ,  $\xi(1-\xi)^2$ ,  $\xi(1-\xi)^2$  and  $\xi(1-\xi)^2$ , where

$$\xi = \frac{2q(1-q)}{1-sq^2}$$

is the sampling frequency of a heterozygote (*Aa*) from the population (Lee *et al.*, 1994). These seven combinations lead to the following deleterious allele frequencies in the female and male gamete pools for the half-sib mating: ( $\frac{1}{3}, \frac{1}{3}$ ), ( $\frac{1}{4}, \frac{1}{3}$ ), ( $\frac{1}{3}, \frac{1}{4}$ ), ( $\frac{1}{4}, \frac{1}{4}$ ), ( $\frac{1}{4}, \frac{1}{4}$ ), (0,  $\frac{1}{4}$ ) and ( $\frac{1}{4}$ , 0). From these values the distribution of the survival of the offspring can be derived. This second distribution is a binomial with parameter *Q*, where

$$Q = 1 - (8/9)^{m_1}(11/12)^{m_2+m_3}(15/16)^{m_4+m_5}1^{m_6+m_7}. \quad (3)$$

To obtain information about the parameters (and the number of lethal equivalents), we calculate the conditional expectations of the proportion of inviable offspring among half-sib families and then marginalize this over the space of missing genotypes by conditioning arguments. The marginal expectation of the proportion of deaths in the *j*th family,

$$\hat{Q}_j = \frac{d_j}{n_j},$$

can then be expressed as a function of *E*:

$$E[\hat{Q}_j] = 1 - (1 - \frac{1}{16}\xi - \frac{5}{48}\xi^2 - \frac{3}{54}\xi^3)^M \approx 1 - e^{-E/16}, \quad (4)$$

where *n<sub>j</sub>* is the number of the planted offspring in the *j*th family and *d<sub>j</sub>* is the number of unviable offspring among them. Similarly, the marginal variance of  $\hat{Q}_j$  is approximated as  $(e^{-E/16} - e^{-31E/256})/n_j + e^{-31E/256} - e^{-E/8}$  (refer to Lee *et al.* (1996) for the procedure of deriving these marginal expectations).

We may apply this approach directly in extending Lee *et al.* (1996) to more complex mating systems. However, we note that, in such complex mating designs, the number of possible combinations to be considered in the sampling distribution of the parents can rapidly become too numerous for the derivation of marginal moments to be analytically tractable. Furthermore, among the combinations to be considered in the sampling of the parents, some have a disproportionate effect on the estimates, in part because of their higher frequency in the population. In the half-sib case, the major contributor is the *m*<sub>5</sub> combination because only the common parent is heterozygous in this case, which is the most frequent and prevailing type of presence of deleterious alleles over all the other types. That is, from the survival distribution in eqn 3, the mortality *Q* from only *m*<sub>5</sub> can be approximated by  $1 - (\frac{15}{16})^{m_5}$ .

Therefore, under the assumption of low deleterious allele frequencies, a good approximation of the marginal moments can be obtained once we know (i) the frequency of the major combination in the

population, and (ii) the deleterious allele frequencies in the male and female gamete gene pools deriving from it. In the half-sib case, these are  $\xi(1-\xi)$ ,  $1/4$  and  $1/4$ , so that we can again derive the approximate mortality of  $Q$  in eqn 3 by  $Q = 1 - (1 - (1/4)^2)^{m_5} = 1 - (\frac{15}{16})^{m_5}$ , where  $m_5$  follows a binomial distribution with parameters  $M$  and  $\xi(1-\xi)$ , considering only the sampling of the prevailing parent resulting in deleterious homozygous offspring. These frequencies for other mating systems are given in Table 1.

Assuming that the heterozygous individual frequency of deleterious alleles ( $\xi$ ) is very low (usually, less than 0.01), we can obtain the asymptotic estimates of  $E$  from the mating designs in Table 1 using standard statistical transformation techniques. Lee *et al.* (1996) show that these estimates are consistent with the true value of  $E$ . For the half-sib mating (stepbrother and stepsister mating) the estimate adjusted by outcrossing (mating between unrelated individuals) for environmental effects on mortality is:

$$\hat{E}_{\text{half-sib}} = -16 \log(1 - \hat{Q}_{\text{half-sib}}) + 16 \log(1 - \hat{Q}_{\text{outcross}}), \quad (5)$$

where  $\hat{Q}_{\text{half-sib}}$ ,  $\hat{Q}_{\text{outcross}}$  are the sample means of family mortality probabilities for the half-sib mating and outcrossing experiments, respectively. The variance of the half-sib estimate (5) can be approximated as

$$\text{var}(\hat{E}) \approx 256 \left[ \left( \frac{e^{E/16}}{1-p_e} - e^{E/256} \right) / \bar{n}_h + e^{E/256} - 1 \right] / N_H + 256\sigma_e^2, \quad (6)$$

where  $p_e$  is the mortality proportion for environmental effects,

$$\sigma_e^2 = \left( \frac{p_e}{(1-p_e)\bar{n}_o} \right) / N_O$$

is the variance of mortality attributable to environmental effects,  $\bar{n}_h$ ,  $\bar{n}_o$  are the harmonic means of the numbers of offspring, and  $N_H$ ,  $N_O$  are the numbers of families in half-sib and outcrossing experiments, respectively. Formulae for the estimates of  $E$  and their variances for the mating designs described in Table 1 are given in Table 2.

### Extension to a mixture of mating systems

In human populations, generally one has to consider different types of mating simultaneously. Considering different matings as independent events, i.e. there is no common ancestry, a linear combination of all separate estimates can be used to obtain a global estimate of the number of lethal equivalents:

$$\hat{E}_{\text{mix}} = w_1 \hat{E}_1 + \dots + w_k \hat{E}_k, \quad (7)$$

where  $\hat{E}_i$ s are separate estimates for the  $k$  different types of mating designs and  $w_i$ s are constant weights such that  $w_1 + \dots + w_k = 1$  and  $w_i \geq 0$  for all  $i = 1, \dots, k$ . Note that these  $k$  distinct estimates have the common mean  $E$  and that variances  $c_i$  can be expressed as in Table 2. The weights  $w_i$ s are obtained by, for instance, minimizing the variance of  $\hat{E}_{\text{mix}}$ . In this case, using the method of Lagrange multipliers, the solution of each  $w_i$  under the constraints described above is calculated as

$$\hat{w}_i = kc_i^{-1} / \bar{c}, \quad (8)$$

where  $\bar{c}$  is the harmonic mean of  $c_i$ s, i.e.  $\bar{c} = k / \sum c_i^{-1}$ . Then, the variance of the mixture estimate is  $\text{var}(\hat{E}_{\text{mix}}) = 1 / \sum c_i^{-1} = \bar{c} / k$ . We note that other (statis-

**Table 1** Various mating designs: inbreeding coefficient, parental frequency of the major combination resulting in deleterious homozygous offspring, and deleterious gene frequencies in gene pools of progeny

| Design        | $F$  | Freq. of heterozygotes in a parent | Gene pool |        |               |
|---------------|------|------------------------------------|-----------|--------|---------------|
|               |      |                                    | Male      | Female | Male × female |
| Selfing       | 1/2  | $\xi$                              | 1/2       | 1/2    | 1/4           |
| Back-cross    | 1/4  | $\xi(1-\xi)$                       | 1/2       | 1/4    | 1/8           |
| Full-sib      | 1/4  | $2\xi(1-\xi)$                      | 1/4       | 1/4    | 1/16          |
| Half-sib      | 1/8  | $\xi(1-\xi)$                       | 1/4       | 1/4    | 1/16          |
| Uncle-niece   | 1/8  | $2\xi(1-\xi)$                      | 1/4       | 1/8    | 1/32          |
| First-cousin  | 1/16 | $2\xi(1-\xi)$                      | 1/8       | 1/8    | 1/64          |
| 1½-cousin     | 1/32 | $2\xi(1-\xi)$                      | 1/8       | 1/16   | 1/128         |
| Second-cousin | 1/64 | $2\xi(1-\xi)$                      | 1/16      | 1/16   | 1/256         |

**Table 2** Asymptotic estimates of lethal equivalents and their variances

| Design                 | $\hat{E}$                | Var( $\hat{E}$ )  |
|------------------------|--------------------------|---|
| Selfing                | $-4 \log(1 - \hat{Q})$   | $16 \left[ \left( \frac{e^{E/4}}{1 - p_e} - e^{E/16} \right) / \bar{n} + e^{E/16} - 1 \right] / N + 16\sigma_e^2$               |
| Back-cross             | $-8 \log(1 - \hat{Q})$   | $64 \left[ \left( \frac{e^{E/8}}{1 - p_e} - e^{E/16} \right) / \bar{n} + e^{E/16} - 1 \right] / N + 64\sigma_e^2$               |
| Full-sib               | $-8 \log(1 - \hat{Q})$   | $64 \left[ \left( \frac{e^{E/8}}{1 - p_e} - e^{E/64} \right) / \bar{n} + e^{E/64} - 1 \right] / N + 64\sigma_e^2$               |
| Half-sib               | $-16 \log(1 - \hat{Q})$  | $256 \left[ \left( \frac{e^{E/16}}{1 - p_e} - e^{E/256} \right) / \bar{n} + e^{E/256} - 1 \right] / N + 256\sigma_e^2$          |
| Uncle-niece            | $-16 \log(1 - \hat{Q})$  | $256 \left[ \left( \frac{e^{E/16}}{1 - p_e} - e^{E/512} \right) / \bar{n} + e^{E/512} - 1 \right] / N + 256\sigma_e^2$          |
| First-cousin           | $-32 \log(1 - \hat{Q})$  | $1024 \left[ \left( \frac{e^{E/32}}{1 - p_e} - e^{E/2048} \right) / \bar{n} + e^{E/2048} - 1 \right] / N + 1024\sigma_e^2$      |
| $1\frac{1}{2}$ -cousin | $-64 \log(1 - \hat{Q})$  | $4096 \left[ \left( \frac{e^{E/64}}{1 - p_e} - e^{E/8192} \right) / \bar{n} + e^{E/8192} - 1 \right] / N + 4096\sigma_e^2$      |
| Second cousin          | $-128 \log(1 - \hat{Q})$ | $16384 \left[ \left( \frac{e^{E/128}}{1 - p_e} - e^{E/32768} \right) / \bar{n} + e^{E/32768} - 1 \right] / N + 16384\sigma_e^2$ |

Note:  $\hat{Q} = \frac{1}{N} \sum \hat{Q}_i$ ,  $N$  = no. of families,  $\bar{n}$  = harmonic mean of the no. of offspring of the  $N$  families,  $\sigma_e^2$  = variance attributable to environmental effects.

tically) desirable estimates such as maximum likelihood estimators of the  $w_i$  are not feasible because each estimate  $\hat{E}_i$  was derived by the *marginalized methods of moments*, summing out the unobserved genotype variables and then evaluating the expectations (Lee *et al.*, 1996).

**Application: how good are the MCM estimates?**

We analysed five data sets from the literature using the previous method. Four of them are from Morton *et al.* (1956), and the fifth is from Rao & Inbaraj (1977). Because we do not know the family structures, we shall assume that two different births are from two distinct nonconsanguineous families. The first two data sets, Morbihan and Loir et Cher, are from Catholic marriage dispensations issued during 1919–25 in two French departments; the following two (Arner and Bemiss) are from infantile deaths

from consanguineous marriages in early American populations; the last one (Tamil Nadu) is from gestation and infantile mortality data in Tamil Nadu, South India (Table 3).

Estimates of the number of lethal equivalents and their corresponding variances are given in Tables 4 and 5. The pooled estimate of  $E$  is the weighted average of separate estimates of  $E$ , and the pooled variance (second last column) is the harmonic mean of the separate estimates divided by the number of different mating types in each of the corresponding data. These are derived as follows. Each separate estimate and its variance can be obtained directly by applying the formulae in Table 2. From these separate variance estimates, we calculate weights  $\hat{w}_i$  as described in the previous section. The pooled estimate of  $E$  is the weighted average of separate estimates of  $E$  from several types of consanguineous marriage data. For example, in Table 4 the pooled estimates of  $E$  with respect to stillbirth and neonatal

deaths are 2.17 and 1.17 for the two French populations; for infantile and juvenile deaths, these are 2.75 and 1.93, and for the total mortality they are 5.13 and 2.87. The pooled variance is the harmonic mean of the separate variance estimates divided by the number of different types of mating designs (intrarelation marriage types). The separate estimates of  $E$  from different marriage types of a population are often found to be quite different. Usually, the closest-relation mating design, first-cousin marriage in Table 4, leads to the best precision for

the estimation. Estimates of  $1\frac{1}{2}$ -cousin data are worse than those of second-cousin data because the total number of observations for the  $1\frac{1}{2}$ -cousin data is smaller than for the latter. However, our pooled estimate, by construction, will have a smaller variance than any separate estimates. We note that our point estimates of  $E$  are very close to those of Morton *et al.* (1956) derived by a weighted regression of the number of survivors on the inbreeding coefficient ( $F$ ); our estimates are twice the MCM estimates because we consider  $E$  'per individual'

**Table 3** Consanguineous marriage and infantile mortality data from Morton *et al.* (1956) and Rao & Inbaraj (1977)

| Population                | Uncle-niece | First-cousin | $1\frac{1}{2}$ -cousin | Second-cousin | Unrelated |
|---------------------------|-------------|--------------|------------------------|---------------|-----------|
| Morbihan                  |             |              |                        |               |           |
| Neonatal deaths           |             | 51/461       | 3/78                   | 23/309        | 72/1628   |
| Juvenile deaths           |             | 64/410       | 17/75                  | 32/286        | 138/1556  |
| Total                     |             | 115/461      | 20/78                  | 55/309        | 210/628   |
| Loir et Cher              |             |              |                        |               |           |
| Neonatal deaths           |             | 18/282       | 6/105                  | 11/240        | 36/1117   |
| Juvenile deaths           |             | 32/264       | 1/99                   | 17/229        | 60/1081   |
| Total                     |             | 50/282       | 7/105                  | 28/240        | 96/1117   |
| Arner                     |             |              |                        |               |           |
| Deaths under age 20 years |             | 113/672      |                        | 211/1471      | 370/3184  |
| Bemiss                    |             |              |                        |               |           |
| Dying young               | 78/207      | 637/2778     |                        | 85/513        | 134/837   |
| Tamil Nadu                |             |              |                        |               |           |
| Neonatal (rural)          | 307/1308    | 430/1991     |                        |               | 939/4449  |
| Neonatal (urban)          | 83/371      | 219/989      |                        |               | 884/4251  |
| Total                     | 390/1679    | 649/2980     |                        |               | 1823/8700 |

**Table 4** Marginalized moments estimates of the number of lethal equivalents, their 95% confidence intervals and twice the MCM estimates for the Morbihan and Loir et Cher populations (under the assumption that each individual was obtained from a different pair of parents)

| Population      |                             | Separate estimates |                        |                 | Pooled estimate | $2 \times$ MCM estimate |
|-----------------|-----------------------------|--------------------|------------------------|-----------------|-----------------|-------------------------|
|                 |                             | First-cousin       | $1\frac{1}{2}$ -cousin | Second-cousin   |                 |                         |
| Morbihan        |                             |                    |                        |                 |                 |                         |
| Neonatal deaths | $\hat{E} \pm 2\hat{\sigma}$ | $2.30 \pm 1.10$    | $0 \pm 3.19$           | $4.11 \pm 4.35$ | $2.17 \pm 1.02$ | 2.25                    |
|                 | $\hat{w}_i$                 | 0.844              | 0.101                  | 0.055           |                 |                         |
| Juvenile deaths | $\hat{E} \pm 2\hat{\sigma}$ | $2.46 \pm 1.45$    | $10.51 \pm 8.06$       | $3.30 \pm 5.74$ | $2.75 \pm 1.39$ | 2.86                    |
|                 | $\hat{w}_i$                 | 0.912              | 0.030                  | 0.058           |                 |                         |
| Total           | $\hat{E} \pm 2\hat{\sigma}$ | $4.76 \pm 1.82$    | $10.12 \pm 8.60$       | $7.41 \pm 7.20$ | $5.13 \pm 1.73$ | 5.11                    |
|                 | $\hat{w}_i$                 | 0.902              | 0.040                  | 0.058           |                 |                         |
| Loir et Cher    |                             |                    |                        |                 |                 |                         |
| Neonatal deaths | $\hat{E} \pm 2\hat{\sigma}$ | $1.06 \pm 1.05$    | $1.67 \pm 3.16$        | $1.81 \pm 3.88$ | $1.17 \pm 0.97$ | 1.15                    |
|                 | $\hat{w}_i$                 | 0.843              | 0.095                  | 0.062           |                 |                         |
| Juvenile deaths | $\hat{E} \pm 2\hat{\sigma}$ | $2.31 \pm 1.54$    | $0 \pm 3.26$           | $2.56 \pm 5.16$ | $1.93 \pm 1.34$ | 1.82                    |
|                 | $\hat{w}_i$                 | 0.763              | 0.170                  | 0.067           |                 |                         |
| Total           | $\hat{E} \pm 2\hat{\sigma}$ | $3.37 \pm 1.86$    | $0 \pm 4.00$           | $4.38 \pm 6.45$ | $2.87 \pm 1.63$ | 2.96                    |
|                 | $\hat{w}_i$                 | 0.769              | 0.167                  | 0.064           |                 |                         |

instead of the 'per gamete' they used.

We also analyse two data sets for early American populations, Arner and Bemiss, which reported infantile deaths from consanguineous marriages. Because of uncertainty in the collection of these data, we exclude some questionable parts of these data sets such as mating data farther related than second-cousin. Our estimates of  $E$  are 2.13 and 3.30 for the Arner and Bemiss populations, respectively. Finally, we analyse the mortality data from consanguineous marriages of a population in South India (Rao & Inbaraj, 1977); the mortality was observed for both gestation and infantile periods. Compared with the preceding examples, our estimates of  $E$  for these two, both rural and urban populations, are very low. Because these populations have very high rates of consanguineous marriages (not farther than first-cousin marriages, 40.6%), most deleterious genes in these populations may have been purged as suggested by Kang *et al.* (1994). Furthermore, the death rate for the unrelated group of this case is very high compared with those of the other cases, so many genetic deaths may be confounded with the environmental deaths. For comparison, we also show the MCM estimates from the literature (last column, Tables 4 and 5).

**Conclusions**

There is little use in having an estimate of a quantity unless one can ask questions about it (Lewontin, 1966). Estimates of the number of lethal equivalents

were often given alone (for example, Morton, 1975; Vogel & Motulsky, 1982) or accompanied by  $R^2$  or  $\chi^2$ -tests (Morton, 1960; Rao & Inbaraj, 1977; Ralls *et al.*, 1988) because no direct estimation method of the number of lethal equivalents was available. The method presented here is a first step to remedy this. Of course, because of the inherent complexity of the problem, the method also has its limitations. Notably, to minimize the calculations, we have resorted to an asymptotic estimation procedure. Consequently, a certain number of observations (families) are needed to reach asymptotic convergence. In general, this number has to be greater than 30 (Lee *et al.*, 1996).

Our point estimates of  $E$  are mostly similar to those obtained through the MCM method. This is hardly surprising for two reasons: (i) the assumptions are basically the same for both methods, and (ii) in the computation of our estimates, we assume that each individual is sampled from a distinct family, so that our evaluation of each separate estimate is equivalent to that of MCM. The variances attached to each global estimate are fairly small, although some of the separate estimates can be rather large, as for  $1\frac{1}{2}$ -cousins in the first data set. The latter was almost certainly the result of the small number of individuals for this level of inbreeding, but generally, and as one would expect, the higher the inbreeding level, the lower the variance. Even if the pooled variance is small, we should note that because we used a linear combination of separate estimates minimizing the combined variance,

**Table 5** Marginalized moments estimates of the number of lethal equivalents, their 95% confidence intervals and twice the MCM estimates for the Arner, Bemiss, and Tamil Nadu populations (under the assumption that each individual was obtained from a different pair of parents)

| Population                |  | Separate estimates   |                       |                      | Pooled estimate | 2 × MCM estimate |
|---------------------------|--|----------------------|-----------------------|----------------------|-----------------|------------------|
|                           |  | Uncle–niece          | First-cousin          | Second-cousin        |                 |                  |
| Arner                     |  |                      |                       |                      |                 |                  |
| Deaths under age 20 years | $\hat{E} \pm 2\hat{\sigma}$<br>$\hat{w}_i$ |                      | 1.94 ± 1.18<br>0.658  | 2.41 ± 1.64<br>0.342 | 2.13 ± 0.96     | 2.06             |
| Bemiss                    |  |                      |                       |                      |                 |                  |
| Dying young               | $\hat{E} \pm 2\hat{\sigma}$<br>$\hat{w}_i$ | 4.78 ± 1.79<br>0.292 | 2.75 ± 1.17<br>0.685  | 0.86 ± 6.35<br>0.023 | 3.30 ± 0.97     | 3.46             |
| Tamil Nadu                |  |                      |                       |                      |                 |                  |
| Before age 1 year (rural) | $\hat{E} \pm 2\hat{\sigma}$<br>$\hat{w}_i$ | 0.49 ± 0.55<br>0.729 | 0.20 ± 0.90<br>0.271  |                      | 0.41 ± 0.47     | 0.40             |
| Before age 1 year (urban) | $\hat{E} \pm 2\hat{\sigma}$<br>$\hat{w}_i$ | 0.32 ± 0.93<br>0.625 | 0.55 ± 1.20<br>0.375  |                      | 0.41 ± 0.73     | 0.41             |
| Total                     | $\hat{E} \pm 2\hat{\sigma}$<br>$\hat{w}_i$ | 0.47 ± 0.46<br>0.702 | 0.336 ± 0.71<br>0.298 |                      | 0.428 ± 0.39    | *                |

our combined estimate is greatly affected by the separate estimate having the smallest variance, usually the estimate from the closest-relation mating design. We also note that in our evaluation of Tables 4 and 5, we assumed that all individuals were from distinct families (because we could not obtain all the family structure of the data). Thus, we may be underestimating these variances as equations given in Table 2 indicate that increasing the number of families is more effective in minimizing the variance than increasing the number of offspring in each family. Because of these restrictions, applying our approach directly to the human data sets previously obtained may not be completely valid; but designing a study fulfilling the requirements of the method should not cause any particular problem. On the contrary, because the method does not necessarily require different levels of inbreeding, estimating the number of lethal equivalents becomes easier. Finally, as noted in Lee *et al.* (1996), our approach provides consistent estimates of  $E$  and their variances can be expressed as functions of design factors, number of families and number of offspring in each family, so that we can control the precision of our estimation by varying these experimental factors.

### Acknowledgements

We thank an anonymous reviewer for helpful comments on the manuscript and Professors J. F. Crow and C. Denniston for useful discussions. This study was supported by both the Swedish National Energy Commission (NUTEK) and NSF grants DMS 90-05833, BIR 95-04393 in the USA.

### References

- BISHIR, J. AND NAMKOONG, G. 1987. Unsound seeds in conifers: estimation of number of lethal alleles and of magnitudes of effects associated with the maternal parent. *Silvae Genet.*, **36**, 180–185.
- CAVALLI-SFORZA, L. L. AND BODMER, W. F. 1971. *The Genetics of Human Populations*, W. H. Freeman, San Francisco.
- CHAKRABORTY, R. AND CHAKRAVARTI, A. 1977. On consanguineous marriages and the genetic load. *Hum. Genet.*, **36**, 47–54.
- CROW, J. F. 1993. Mutation, mean fitness and genetic load. In: Futuyma, D. and Antonovics, J. (eds) *Oxford Surveys in Evolutionary Biology*, **9**, pp. 3–42. Oxford University Press, New York.
- EWENS, W. J. 1993. Beanbag genetics and after. In: Majumder, P. P. (ed.) *Human Population Genetics*. Plenum Press, New York.
- GILLESPIE, J. H. 1991. The burden of genetic load. *Science*, **254**, 1049–1951.
- KANG, H., HARDNER, C. AND GULLBERG, U. 1992. Lethal equivalents in Willow, *Salix viminalis*. *Silvae Genet.*, **41**, 110–117.
- KANG, H., LEE, J. K. AND NORDHEIM, E. V. 1994. Number of lethal loci and lethal equivalents in willow, *Salix viminalis*. *Silvae Genet.*, **43**, 138–145.
- KERR, W. E. 1975. Population genetic studies in bees (Apidae, Hymenoptera). I. Genetic load (*Apis mellifera*). *An. Acad. Bras. Cienc.*, **47**, 319–334.
- KHLAT, M. 1988. Consanguineous marriage and reproduction in Beirut, Lebanon. *Am. J. Hum. Genet.*, **43**, 188–196.
- KLEKOWSKI, E. J. 1988. *Mutation, Developmental Selection, and Plant Evolution*. Columbia University Press, New York.
- LEE, J. K., NEWTON, M. A., NORDHEIM, E. V. AND KANG, H. 1994. Inference on lethal gene studies via Bayesian Markov chain simulation. *Comp. Sci. Stat.: Proc. of the 26th Symp. on the Interface*, pp. 410–414.
- LEE, J. K., NORDHEIM, E. V. AND KANG, H. 1996. Inference for lethal gene estimation with application in plants. *Biometrics*, **52**, (in press).
- LEWONTIN, R. C. 1966. On the measurement of relative viability. *Syst. Zool.*, **15**, 141–142.
- MAKOV, E. AND BITTLES, A. H. 1986. On the choice of mathematical models for the estimates of lethal gene equivalents in man. *Heredity*, **69**, 308–314.
- MORTON, N. E. 1960. The mutation load due to detrimental genes in man. *Am. J. Hum. Genet.*, **12**, 348–364.
- MORTON, N. E. 1975. Kinship, fitness and evolution. In: Salzano, F. M. (ed.) *The Role of Natural Selection in Human Evolution*, pp. 133–154. North-Holland, Amsterdam.
- MORTON, N. E., CROW, J. F. AND MULLER, H. J. 1956. An estimation of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. U.S.A.*, **42**, 855–863.
- RALLS, K., BALLOU, J. D. AND TEMPLETON, A. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conserv. Biol.*, **2**, 185–193.
- RAO, P. S. S. AND INBARAJ, S. G. 1977. Inbreeding effects on human reproduction in Tamil Nadu of South India. *Ann. hum. Genet.*, **41**, 87–98.
- SAVOLAINEN, O., KÄRKKÄINEN, K. AND KUITTINEN, H. 1992. Estimating numbers of embryonic lethals in conifers. *Heredity*, **69**, 308–314.
- SORENSEN, F. 1969. Embryonic genetic load in coastal Douglas-fir, *Pseudotsuga menziesii* var. *menziesii*. *Am. Nat.*, **103**, 389–398.
- VOGEL, F. AND MOTULSKY, A. G. 1982. *Human Genetics. Problems and Approaches*. Springer, Berlin.
- WALLACE, B. 1991. *Fifty Years of Genetic Load. An Odyssey*. Cornell University Press, New York.