Heredity **57** (1986) 377-380

Received 19 February 1986

On the choice of mathematical models for the estimation of lethal gene equivalents in man

E. Makov* and A. H. Bittles†

* Department of Computer Science and Statistics, Queen Mary College, Mile End Road, London E14 NS, U.K. and † Department of Anatomy and Human Biology, King's College, Strand, London WC2R 2LS, U.K.

A range of mathematical models and error distributions was used to examine the validity of linear regression methods for the calculation of lethal gene equivalents. Because of the restricted span of inbreeding coefficient F values available in human studies and the limited number of data points, equivalent results were obtained with all combinations tested. It was concluded that linear regressions should be employed only for the detection of significant inbreeding effects in man and that their application to the estimation of lethal gene equivalents was not warranted.

INTRODUCTION

The measurement of inbreeding depression is of interest both from a general biological viewpoint and in relation to the genetic load carried by a particular group or species. As a means of calculating the number of deleterious recessive mutants carried by a human population, Morton *et al.* (1956) derived the formula:

$$S = e^{-A-BF}$$

where S is the proportion of survivors in a population, A is deaths expressed under random mating, B is deaths arising from the expression of recessive genes via inbreeding and F, the coefficient of inbreeding, is that fraction of gene loci homozygous as a result of consanguinity.

Adoption of the formula was justified on the grounds of presumed independence between genetic and environmental influences with respect to the proportion of survivors and, in order to estimate A and B, S was transformed to reveal the linear relationship:

$$-\ln S = A + BF$$

Under the implicit assumption of normally distributed errors, A and B then could be estimated using linear regression or, as subsequently was suggested (Smith, 1967, 1969), by weighted regression. Reservations have been expressed on the validity of results thus obtained, especially with respect to the additional calculation of B/A ratios for assessing the relative roles of mutation and segregation in the total genetic load (Cavalli-Sforza and Bodmer, 1971; Spiess, 1977). Nevertheless, the method has been extensively applied to the estimation of lethal equivalent genes in a wide range of human populations, for example, Schull (1958), Neel and Schull (1962), Yamaguchi *et al.* (1970), Chakraborty and Chakravarti (1977), Rao and Inbaraj (1979), Azevedo *et al.* (1980) and Freire-Maia (1984).

In the light of recent observations on theoretical and practical limitations to the use of linear regressions for the calculation of lethal gene equivalents (Bittles and Makov, 1985), the aim of the present study was to re-examine the basic statistical assumptions made in the original method (Morton *et al.* 1956).

MATERIALS AND METHODS

Five mathematical models and four error distributions were employed to analyse data from two extensively cited human inbreeding studies (Bemiss, 1858; Sutter and Tabah, 1952, 1953). The functions, linearising transformations and error distributions investigated are summarised in table 1.

* *Present address*: Department of Statistics, University of Haifa, Mount Carmel, Haifa, Israel 31999.
† To whom all correspondence should be addressed.

Table 1 The models, linearising functions and error distributions used in the investigation of human inbreeding effects

Model	Transformation	Error distribution
$M1$: $S = e^{A+BF}$	$\ln S = A + BF$	normal, Poisson, gamma
$M2$: $S = \frac{e^{A+BF}}{1+e^{A+BF}}$	$\ln\left(\frac{S}{1-S}\right) = A + BF$	binomial, normal
$M3$: $S = 1 - e^{-e^{A+BF}}$	$\ln[-\ln(1-S)] = A + BF$	binomial
$M4$: $S = A + BF$	$S = A + BF$	normal
$M5$: $S = \frac{1}{A+BF}$	$\frac{1}{S} = A + BF$	normal, Poisson, gamma

$M1$ was used as, with a change of signs, it is the model which predominantly has been utilised in the literature.

$M2$ and $M3$ were introduced as they are two of the models commonly proposed for estimating proportions (Dobson, 1983).

$M4$ and $M5$ were tested primarily for comparative purposes; their formal adoption could not be justified on *a priori* grounds.

The statistical package GLIM (Generalized Linear Interactive Modelling: Numerical Algorithms Group, Oxford, England) was employed in fitting the 10 combinations of models and error distributions. The parameters A and B were estimated by the method of Maximum Likelihood, iterated with weights corresponding to the different sample sizes until the change in the relative reduction of deviance reached 0·01 (McCullagh and Nelder, 1983). Weights were not utilised for $M2$ and $M3$ with the binomial distribution as, in those cases, the sample size explicitly was incorporated into the modes. The significance of a fit was examined by the F-test, which has been suggested as more appropriate for this purpose than the X^2 test (Baker and Nelder, 1978); the level of significance chosen was 5 per cent.

RESULTS

Applying the ten model/error distribution combinations to the data of Sutter and Tabah (1952, 1953), the most striking general finding was that with reference to the significance ($\alpha = 0{\cdot}05$) or otherwise of inbreeding on the proportion of survivors, all the models and associated error distributions tested produced the same basic conclusions, table 2. The results uniformly indicated that for both age groups in Morbihan and for infantile and juvenile deaths in Loir et Cher, inbreeding had no significant effect on the proportion of survivors. Thus, in these cases, any attempted interpretation of estimates of B and A or B/A ratios essentially would be meaningless. By comparison, all models and error distributions showed that stillbirths and neonatal deaths in Loir et Cher were significantly related to inbreeding. However, it also can be seen that despite the virtual unanimity between the models/error distributions with respect to p values, the estimates for B and A obtained with each combination varied widely, table 3.

From these findings two inter-related questions arise: which of the estimates best describes the relative contributions of genetic and environ-

Table 2 The effects of inbreeding on ante- and post-natal mortality (after Sutter and Tabah 1952, 1953)

		Morbihan		Loir et Cher	
Model	Error distribution	Stillbirths and neonatal deaths	Infantile and juvenile deaths	Stillbirths and neonatal deaths	Infantile and juvenile deaths
$M1$	normal	NS	NS	S $0{\cdot}01 < p < 0{\cdot}025$	NS
$M1$	Poisson	NS	NS	S $0{\cdot}01 < p < 0{\cdot}025$	NS
$M1$	gamma	NS	NS	S $0{\cdot}01 < p < 0{\cdot}025$	NS
$M2$	binomial	NS	NS	S $0{\cdot}025 < p < 0{\cdot}05$	NS
$M2$	normal	NS	NS	S $0{\cdot}025 < p < 0{\cdot}05$	NS
$M3$	binomial	NS	NS	S $0{\cdot}025 < p < 0{\cdot}05$	NS
$M4$	normal	NS	NS	S $0{\cdot}025 < p < 0{\cdot}05$	NS
$M5$	normal	NS	NS	S $0{\cdot}01 < p < 0{\cdot}025$	NS
$M5$	Poisson	NS	NS	S $0{\cdot}01 < p < 0{\cdot}025$	NS
$M5$	gamma	NS	NS	S $0{\cdot}01 < p < 0{\cdot}025$	NS

NS: non-significant
S: significant

Table 3 Derivation of A and B values for stillbirths and neonatal deaths, Loir et Cher

Model	Error distribution	A	B	B/A
$M1$	normal	−0·03348*	−0·5680*	16·9653
$M1$	Poisson	4·572†	−0·5669	0·1240
$M1$	gamma	−0·0335	−0·5659	16·8925
$M2$	binomial	3·358	−11·56	3·4425
$M2$	normal	3·377	−11·32	3·3521
$M3$	binomial	1·223	−3·659	2·9918
$M4$	normal	0·9671	−0·5393	0·5576
$M5$	normal	1·034	0·5983	0·5786
$M5$	Poisson	0·01034†	0·005972	0·5775
$M5$	gamma	1·034	0·5961	0·5765

* These values are negative, unlike the estimates obtained by Morton *et al.*, since logarithmic instead of negative logarithmic transformation has been used.
† For the Poisson error distributions, the proportion of survivors were transformed into counts by multiplying ×100, hence the different A values.

mental factors to ante- and post-natal mortality within individual human populations, and which model/error distribution combination is most applicable to comparative population studies? Unfortunately, for two major statistical reasons, no answers are readily forthcoming as:

(a) In the majority of studies the span of F values is restricted to the interval 0 to 0·0625, or where uncle-niece or double first cousin data are available 0 to 0·125, by comparison with the theoretical total range of 0 to 1.

(b) Only a very limited number of data points is available, usually a maximum of four or, exceptionally, five in number.

The potential distorting effects of both factors on the investigation of data relating to mortality/survivorship is seen by reference to the historical study of Bemiss (1858) on childhood mortality. As analysed by Morton *et al.* (1956) five data points were used, $F = 0$, 0·0039, 0·0156, 0·0625 and 0·125 which, with the ten model/error distribution combinations under test, gave the range of values presented in table 4. However 31 cases of incest ($F = 0{\cdot}25$), in which no deaths were recorded, were omitted from the analysis. If these are restored to the data set rather different values are obtained, table 5. Some models/error distributions increased in significance while others became non-significant. Clearly, the use of six data points and the extended interval of 0 to 0·25 permitted differentiation between a number of the models and error distributions employed. For example, in the Bemiss study the binomial distribution tested in conjunction with M_2 and M_3 confidently could be rejected but this conclusion may not necessarily have general application.

Table 4 The effects of inbreeding on childhood mortality, $F = 0 - 0{\cdot}125$ (after Bemiss, 1858)

Model	Error distribution	Result
$M1$	normal	$S\ 0{\cdot}01 < p < 0{\cdot}025$
$M1$	Poisson	$S\ 0{\cdot}01 < p < 0{\cdot}025$
$M1$	gamma	$S\ 0{\cdot}01 < p < 0{\cdot}025$
$M2$	binomial	$S\ 0{\cdot}001 < p < 0{\cdot}005$
$M2$	normal	$S\ 0{\cdot}001 < p < 0{\cdot}005$
$M3$	binomial	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M4$	normal	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M5$	normal	$S\ 0{\cdot}01 < p < 0{\cdot}025$
$M5$	Poisson	$S\ 0{\cdot}01 < p < 0{\cdot}025$
$M5$	gamma	$S\ 0{\cdot}01 < p < 0{\cdot}025$

Table 5 The effects of inbreeding on childhood mortality, $F = 0 - 0{\cdot}25$ (after Bemiss, 1858)

Model	Error distribution	Result
$M1$	normal	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M1$	Poisson	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M1$	gamma	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M2$	binomial	NS
$M2$	normal	$S\ p < 0.01$
$M3$	binomial	NS
$M4$	normal	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M5$	normal	$S\ 0{\cdot}001 < p < 0{\cdot}005$
$M5$	Poisson	$S\ 0{\cdot}005 < p < 0{\cdot}01$
$M5$	gamma	$S\ 0{\cdot}005 < p < 0{\cdot}01$

CONCLUSION

In view of the data structure, many different models could adequately detect significant inbreeding effects. However, each model typically would result in different values for A, B and B/A. Although only model M_1 commonly has been employed in assessing the effects of inbreeding, there appears to be no compelling reason for its exclusive usage and hence for the uncritical acceptance of B and A values obtained by this method. A more searching analysis of the optimum model/error distribution for the definition of inbreeding effects in man would require even richer data than has been available and studies of the necessary size and range would rarely, if ever, be attainable with human populations. Thus the applicability of the line of investigation originated by Morton *et al.* and followed, in extended version, by the present study effectively is limited to the

detection of significant human inbreeding influence. It should not be extended to attempted quantification of lethal gene equivalents, especially in those cases in which no significant inbreeding effect on survivorship has been demonstrated. Where detailed animal pedigree data is available, for example, in certain breeding stocks or zoo populations, a satisfactory differentiation of the models and their respective error distributions may prove to be feasible. An analysis of this nature would be of considerable general biological interest and additionally could provide valuable guidelines in assessing the results of human studies.

REFERENCES

AZEVEDO, E. S., FREIRE-MAIA, N., AZEVEDO, M. C. C., WEINER, T. A. AND SOUZA, M. M. M. 1980. Inbreeding in a Brazilian general hospital. *Ann. Hum. Genet. 43*, 255–264.

BAKER, R. J. AND NELDER, J. A. 1978. *GLIM Manual*, (release 3), Numerical Algorithms Group, Oxford.

BEMISS, S. M. 1958. Report on influence of marriages of consanguinity upon offspring. *Trans. Amer. Med. Assoc. 11*, 319–425.

BITTLES, A. H. AND MAKOV, E. 1985. Linear regressions in the calculation of lethal gene equivalents in man. *Ann. Hum. Biol. 12*, 287–289.

CAVALLI-SFORZA, L. L. AND BODMER, W. F. 1971. *The Genetics of Human Populations*, W. H. Freeman, San Francisco, pp. 360–365.

CHAKRABORTY, R. AND CHAKRAVARTI, A. 1977. On consanguineous marriage and the genetic load. *Hum. Genet. 36*, 47–54.

DOBSON, A. J. 1983. *Introduction to Statistical Modelling*, Chapman and Hall, London.

FREIRE-MAIA, N. 1984. Effects of consanguineous marriages on morbidity and precocious mortality: genetic counselling. *Amer. J. Med. Genet. 18*, 401–406.

McCULLAGH, P. AND NELDER, J. A. 1983 *Generalized Linear Models*, Chapman and Hall, London.

MORTON, N. E., CROW, J. F. AND MULLER, H. J. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. (U.S.A.) 42*, 855–863.

NEEL, J. V. AND SCHULL, W. J. 1962. The effect of inbreeding on mortality and morbidity in two Japanese cities. *Proc. Natl. Acad. Sci. (U.S.A.) 48*, 573–582.

RAO, P. S. S. AND INBARAJ, S. G. 1979. Trends in human reproductive wastage in relation to long-term practice of inbreeding. *Ann. Hum. Genet. 42*, 401–413.

SCHULL, W. J. 1958. Empirical risks in consanguineous marriages: sex ratio, malformation and viability. *Amer. J. Hum. Genet. 10*, 294–343.

SMITH, C. A. B. 1967. Note on a paper by S. Kumar, R. A. Pai and M. S. Swaminathan. *Ann. Hum. Genet. 31*, 146–147.

SMITH, C. A. B. 1969. Corrigenda. *Ann. Hum. Genet. 32*, 419.

SPIESS, E. B. 1977. *Genes in Populations*, John Wiley, New York, pp. 604–608.

SUTTER, J. AND TABAH, L. 1952. Effets de la consanguinité et de l'endogamie. Une enquete en Morbihan et Loir et Cher. *Population 7*, 249–266.

SUTTER, J. AND TABAH, L. 1953. Structure de la mortalité dans les familles consanguines. *Population 8*, 511–526.

YAMAGUCHI, M. YANASE, T., NAGANO, H. AND NAKAMOTO, N. 1970. Effects of inbreeding on mortality in Fukuoka population. *Amer. J. Hum. Genet. 22*, 145–159.