Appl. Statist. (1988)
37, *No.* 1, *pp.* 87–94

A Method for Computing Profile-Likelihood-Based Confidence Intervals

By D. J. VENZON† and S. H. MOOLGAVKAR

Fred Hutchinson Cancer Research Center, Seattle, USA

[Received April 1987; Revised August 1987]

SUMMARY

The method of constructing confidence regions based on the generalised likelihood ratio statistic is well known for parameter vectors. A similar construction of a confidence interval for a single entry of a vector can be implemented by repeatedly maximising over the other parameters. We present an algorithm for finding these confidence interval endpoints that requires less computation. It employs a modified Newton–Raphson iteration to solve a system of equations that defines the endpoints.

Keywords: Confidence intervals; Profile likelihood

1. 'Profile-Likelihood-Based' Confidence Intervals

A classical method of confidence interval construction is based on the asymptotic normality of the maximum likelihood estimate (m.l.e.) $\hat{\theta}$ of a parameter vector $\theta_0 \in R^k$. It is well known, however, that properties of $\hat{\theta}$ in small samples can be very different from the asymptotic properties. A more robust construction of confidence regions is derived from the asymptotic χ^2 distribution of the generalised likelihood ratio test. Suppose that the log-likelihood $l(\theta)$ is defined for values of θ in a parameter space $\Theta \subset R^k$. An approximate $1-\alpha$ confidence region for θ_0 is $\{\theta \in \Theta : 2[l(\hat{\theta}) - l(\theta)] \leqslant q_k(1-\alpha)\}$, where $q_k(1-\alpha)$ is the $(1-\alpha)$th quantile of the χ^2 distribution on k degrees of freedom.

Confidence intervals for individual entries of θ_0 can be similarly defined, as, for instance, in Cox (1970, p. 88). This 'profile likelihood' method reduces $l(\theta)$ to a function of a single parameter of interest by treating the others as nuisance parameters and maximising over them. Specifically, for $\beta = \theta_j$ the parameter of interest, let us denote the families of parallel hyperplanes in parameter space $\{\theta \in \Theta : \theta_j = \beta\}$ by $\Theta_j(\beta)$. Then

$$\bar{l}_j(\beta) = \max_{\theta \in \Theta_j(\beta)} l(\theta) \tag{1}$$

is called the profile likelihood for β. An approximate $1-\alpha$ profile-likelihood-based confidence interval for $(\theta_0)_j$ is $\{\beta : 2[l(\hat{\theta}) - \bar{l}_j(\beta)] \leqslant q_1(1-\alpha)\}$. We will write $\omega = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)^t$ for the vector of remaining parameters, and $\hat{\omega}(\beta)$ for the value of ω that achieves the maximum in equation (1). We assume that $\hat{\omega}(\beta)$ lies in the interior of $\Theta_j(\beta)$.

† *Address for correspondence*: Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, USA.

 0035–9254/88/37087 $2.00

It appears that little or no attention has been given in the statistical literature to efficient computation of the endpoints of profile-likelihood-based confidence intervals. Presumably the technique derived directly from the definition is most often employed. This would compute the endpoints of the jth confidence interval by finding the zeroes of the function of a single variable $l_j(\beta) - l^*$, where the constant $l^* = l(\hat{\theta}) - [q_1(1-\alpha)]/2$. Each solution would generally require several iterations with different values of β, and, for each β, the iterative maximisation over $k-1$ parameters to determine $\hat{\omega}(\beta)$. If confidence intervals for each parameter are required, $2k$ such zeroes must be found. The computational effort that this entire procedure requires is at least one of the reasons for its limited use at this time.

The purpose of this paper is to point out that the confidence interval endpoints can also be found as the values of θ_j in the solutions of the system of k equations

$$\begin{bmatrix} l(\theta) - l^* \\ \dfrac{\partial l}{\partial \omega}(\theta) \end{bmatrix} = 0. \tag{2}$$

Numerical solution of this system does not require that intermediate values of the parameter vector lie on $\hat{\omega}(\beta)$ at each iteration. This generally leads to computational savings over the method which alternates between steps in ω and in β. The system (2) also has a familiar form: it differs only in the first equation from the system that defines $\hat{\theta}$. The accurate computation of each endpoint of a profile-likelihood-based interval should therefore be a problem of about the same difficulty as that of finding the m.l.e. $\hat{\theta}$ from an arbitrary starting point. Moreover, since the first line of (2) involves l rather than one of its derivatives, the routines that compute first and second derivatives of l, needed by a number of maximisation algorithms, can also be used for the numerical solution of (2).

There are two features of this problem that suggest specific modifications of the Newton–Raphson algorithm for its solution. These are discussed in the next section. The performance of this method in some simulations and two applications to published data are reported in Sections 3 and 4, respectively. Convergence properties and extensions of the method are taken up in the final section.

2. Basis of the Algorithm

Suppose throughout this section that $\beta = \theta_1$. We assume that $\hat{\theta}$ is given as the initialisation of β and ω for each computation of β_L and β_U, the lower and upper confidence interval endpoints, and we repeatedly use a quadratic approximation to $l(\theta)$. This approximation generally is not accurate enough to step directly from $\hat{\theta}$ to the desired solution $\theta^* = (\beta_L, \hat{\omega}(\beta_L)^t)^t$ or $(\beta_U, \hat{\omega}(\beta_U)^t)^t$, however. The basic step of the algorithm is the Newton–Raphson iteration for the solution of equation (2),

$$\begin{bmatrix} \beta^{(i+1)} \\ \omega^{(i+1)} \end{bmatrix} = \begin{bmatrix} \beta^{(i)} \\ \omega^{(i)} \end{bmatrix} - \begin{bmatrix} \dfrac{\partial l}{\partial \beta} & \left(\dfrac{\partial l}{\partial \omega}\right)^t \\ \dfrac{\partial^2 l}{\partial \beta \partial \omega} & \dfrac{\partial^2 l}{\partial \omega^2} \end{bmatrix}^{-1} \begin{bmatrix} l - l^* \\ \dfrac{\partial l}{\partial \omega} \end{bmatrix} \tag{3}$$

where superscripts (i) on β, ω and θ denote values at the ith iteration, and l and its derivatives on the right-hand side are evaluated at $\theta^{(i)}$.

The matrix whose inverse appears in equation (3) is singular at $\hat{\theta}$ because its first row is zero there. Thus, we need to base the first step away from the m.l.e. on some other considerations. It is also possible, and generally beneficial, to compute a correction to each Newton–Raphson step using the second derivatives of l that are ignored in the first row of (3). These two will now be discussed more fully.

First step: The set of points $\{(\beta, \hat{\omega}(\beta)^t)^t\}$ describes a curve in Θ. (Some examples appear in Figs 1 and 2, to be considered in detail later.) The solution θ^* lies on this curve, at a point where l attains the value l^*. In the absence of other information about the location of the solutions, the indicated direction for the initial step away from $\hat{\theta}$ is along this curve. An approximate technique for this is to take a step in the direction of the tangent to the curve at $\hat{\theta}$, found by differentiating with respect to β. That is,

$$\theta^{(1)} = \begin{bmatrix} \beta^{(1)} \\ \omega^{(1)} \end{bmatrix} = \begin{bmatrix} \hat{\beta} \\ \hat{\omega}(\hat{\beta}) \end{bmatrix} + h \begin{bmatrix} 1 \\ \dfrac{d\hat{\omega}}{d\beta} \end{bmatrix}. \tag{4}$$

To evaluate $d\hat{\omega}/d\beta$, we observe that $\partial l/\partial\omega \equiv 0$ on $\hat{\omega}(\beta)$, and therefore

$$\frac{d}{d\beta}\frac{\partial l}{\partial \omega} = \frac{\partial^2 l}{\partial \omega^2}\frac{d\omega}{d\beta} + \frac{\partial^2 l}{\partial\beta\partial\omega} = 0$$

on $\hat{\omega}(\beta)$. This gives us an expression involving second derivatives of l,

$$\frac{d\hat{\omega}}{d\beta} = -\left(\frac{\partial^2 l}{\partial\omega^2}\right)^{-1}\frac{\partial^2 l}{\partial\beta\partial\omega} = -\begin{bmatrix} \dfrac{\partial^2 l}{\partial\theta_2^2} & \cdots & \dfrac{\partial^2 l}{\partial\theta_k\partial\theta_2} \\ \vdots & & \vdots \\ \dfrac{\partial^2 l}{\partial\theta_2\partial\theta_k} & \cdots & \dfrac{\partial^2 l}{\partial\theta_k^2} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\partial^2 l}{\partial\beta\partial\theta_2} \\ \vdots \\ \dfrac{\partial^2 l}{\partial\beta\partial\theta_k} \end{bmatrix}. \tag{5}$$

The choice of the scale factor h in equation (4) can be based on consideration of the size of the second (Newton–Raphson) step to be taken from the point $\theta^{(1)}$. For instance, we choose h so that if the log-likelihood were exactly quadratic, $\theta^{(1)}$ would fall halfway between $\hat{\theta}$ and the solution θ^*. This implies

$$h = \pm\frac{1}{2}\left[\frac{q_1(1-\alpha)}{-\left\{\dfrac{\partial^2 l}{\partial\beta^2} - \left(\dfrac{\partial^2 l}{\partial\beta\partial\omega}\right)^t\left(\dfrac{\partial^2 l}{\partial\omega^2}\right)^{-1}\dfrac{\partial^2 l}{\partial\beta\partial\omega}\right\}}\right]^{\frac{1}{2}}, \tag{6}$$

where the denominator is evaluated at $\hat{\theta}$. This expression results from taking $\tilde{l}_j(\beta)$ to be quadratic and observing that the denominator is the absolute value of $d^2\tilde{l}_j/d\beta^2$. Other criteria, for instance, involving the determinant of the matrix in equation (3), are possible, but this one seems to work well in practice.

Modified Newton–Raphson algorithm: When begun sufficiently far from $\hat{\theta}$, computation of one confidence interval endpoint can proceed exactly as in (3). We also suggest, however, that all the second derivatives of l be taken into account. This requires only the addition of a correction term along with the standard step. This modification serves two purposes. First, it generally makes convergence much faster. With an exactly quadratic log-likelihood and exact computations, the first modified

Newton–Raphson step would reach the solution from any $\theta^{(1)} \neq \hat{\theta}$. With the more commonly observed log-likelihoods, the correction can be substantial in the first steps, because second derivatives are significant parts of the Taylor series expansion of $l(\theta) - l^*$. An example of this can be seen in Section 4. Secondly, the computations involved help to indicate when the log-likelihood surface is far from quadratic, in which case some technique to prevent instability in the standard Newton–Raphson method might be implemented.

The form of this correction is easily derived. Write $\delta = \theta^* - \theta^{(i)}$ for a fixed iteration number i, and $D(\cdot) = \partial^2 l / \partial \theta^2$. The expansion of the left-side of equation (2) that includes all second derivatives of l is

$$0 = \begin{bmatrix} l(\theta^*) - l^* \\ \dfrac{\partial l}{\partial \omega}(\theta^*) \end{bmatrix} \approx \begin{bmatrix} l(\theta^{(i)}) - l^* \\ \dfrac{\partial l}{\partial \omega}(\theta^{(i)}) \end{bmatrix} + \begin{bmatrix} \dfrac{\partial l}{\partial \theta}(\theta^{(i)}) \\ \dfrac{\partial^2 l}{\partial \theta \partial \omega}(\theta^{(i)}) \end{bmatrix} \delta + \begin{bmatrix} \dfrac{\delta^t D(\theta^{(i)})\delta}{2} \\ 0 \end{bmatrix}.$$

The first two terms on the right lead to the usual iteration (3), but the third term is a complication. Write G for the matrix in the second term and suppress all dependence on $\theta^{(i)}$. Then

$$-G^{-1} \begin{bmatrix} l - l^* \\ \dfrac{\partial l}{\partial \omega} \end{bmatrix} \approx \delta + G^{-1} \begin{bmatrix} \delta^t D \delta / 2 \\ 0 \end{bmatrix} = \delta + (\delta^t D \delta / 2) g_{(1)}, \tag{7}$$

where $g_{(n)}$ is the nth column of G^{-1}. Taking the approximate equality to be exact and writing $-v$ for the left-hand side of (7) and s for the scalar $\delta^t D \delta / 2$, we have $\delta = -v - g_{(1)} s$. This gives us a quadratic equation for s,

$$2s = \delta^t D \delta = g_{(1)}{}^t D g_{(1)} s^2 + 2 v^t D g_{(1)} s + v^t D v. \tag{8}$$

If this equation has two real roots, the one which places

$$\theta^{(i+1)} = \theta^{(i)} - v - g_{(1)} s \tag{9}$$

nearer to $\theta^{(i)}$ is clearly the one to choose. ('Nearer' must be defined in some appropriate distance, e.g., the inner product with the matrix $D(\hat{\theta})$.) If equation (8) has no real roots, then the third derivatives that were set to zero in equation (7) are instead significant. This indicates that a more cautious step than that of equation (3) might be advisable. This could involve a linear search for an optimal step length, for instance.

When the jth parameter, for $j > 1$, is the one of interest, the jth rows and columns of the various vectors and matrices can clearly be used for the computations with only minor changes in the algorithm. A concise listing of the steps of the algorithm is given in the appendix.

3. Performance in Simulations

With the modified Newton–Raphson iteration in equation (9), the algorithm finds a confidence interval endpoint in two steps if $l(\theta)$ is quadratic. We conducted tests of the algorithm on two other types of functions l. Convergence to a solution was defined by the likelihood value being unchanged in successive iterates up to 6 significant digits, and the parameters up to 5 significant digits.

As a first non-trivial case, we applied the method to cubic functions with a local maximum at the origin

$$l(\theta) = -\theta^t\theta(1 + \gamma^t\theta),$$

for constant vectors $\gamma \in R^k$. We generated entries of γ as random uniform deviates on $(-1, 1)$ and then scaled the vector so that both endpoints existed. The algorithm performed well, converging in about 6.3 iterations on average, for several values of k.

As a test with an actual likelihood, we generated random numbers with a Weibull distribution for 20 individuals with binary, log-normal and normal random variables as covariates, and then fit a Weibull regression model (see, for example, Kalbfleisch and Prentice (1980), p. 54). We chose $\alpha = 0.05$, implying $q_1(1-\alpha) = 3.841$. Maximisation of the likelihood with the true parameter value as starting point and using the Newton–Raphson algorithm with the observed second derivative matrix (or the expected second derivative matrix when the first method failed) required about 8 iterations on average for the several parameter values tested. In each case, the number of iterations for finding a confidence interval endpoint was 6 for regression coefficients and 5 or 6 for the shape parameter σ of the Weibull in a large majority of the replicates. The mean numbers were only slightly higher than these.

4. Examples

The application of this method to data taken from Appendix III of Breslow and Day (1980) gives a simple illustration of its use. The data are drawn from a matched case-control study of endometrial cancer. We included two covariates, presence or absence of gallbladder disease and length of estrogen use in months, in the model. These covariates had been recorded for 49 cases matched with four controls each and 8 cases matched with 3 controls each. We fitted an additive relative risk model (see, e.g., Storer *et al.* (1983)), which generally requires much larger sample sizes before asymptotic properties hold, compared to the usual multiplicative relative risk model. The parameter θ is the vector of regression coefficients. Confidence regions for θ are discussed in Moolgavkar and Venzon (1987a).

The curves $\{(\theta_1, \hat{\omega}(\theta_1))^t\}$ and $\{(\hat{\omega}(\theta_2), \theta_2)^t\}$ are shown in Fig. 1, along with the intermediate values in the computations of the 95% profile-likelihood-based confidence intervals for the two coefficients. The intersection of the curves is, of course,

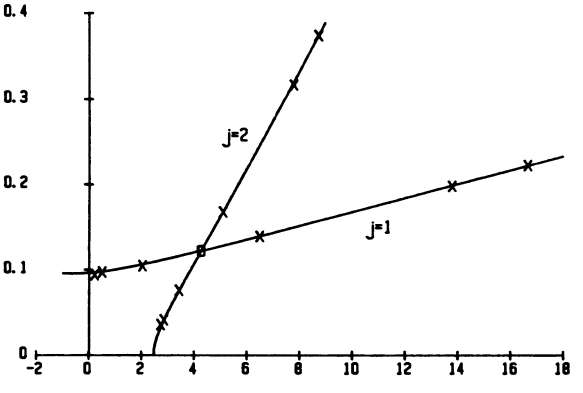

Fig. 1. Solid curves are $\{(\theta_1, \hat{\omega}(\theta_1))^t\}$ (identified as $j = 1$) and $\{(\hat{\omega}(\theta_2), \theta_2)^t\}$ ($j = 2$), 0 gives the location of $\hat{\theta}$, and ×represent points $\theta^{(i)}$ from the computation of confidence interval endpoints, in the example using data from Breslow and Day (1980)

$\hat{\theta} = (4.26, 0.122)^t$. The confidence intervals (0.502, 16.7) and (0.042, 0.374) are not symmetric about $\hat{\theta}_1$ and $\hat{\theta}_2$, but since the curves are nearly linear, the algorithm reaches the endpoints quickly.

A more challenging application comes from the third data set of Williams (1986). It is one of several examples of the estimation of the median lethal dose from 5 replicates at each of 5 dose levels. The model is non-linear in the parameters, with the probability of death p at a dose level x given by $\log[p/(1-p)] = \theta_2(x - \theta_1)$. Since the parameter of interest is θ_1, we show $\{(\theta_1, \hat{\omega}(\theta_1))^t\}$ in Fig. 2. Convergence to the 95% confidence interval endpoints is slower because of the curvature, but eventually attained. The closely spaced iterates near the middle of both sequences are points at which equation (8) had no real roots. At such points we simply took the standard Newton–Raphson step in equation (3) multiplied by 0.1.

The importance of the correction term is demonstrated by the fact that the standard Newton–Raphson algorithm fails to find either endpoint with these data, reaching by the ninth iteration an extreme parameter value where all the probabilities are either zero or one, to machine accuracy. The following table, from the computation of the lower endpoint of the confidence interval, illustrates the contribution of the correction term at each of the early steps shown in Fig. 2:

i	$\theta^{(i)}$	$\theta^{(i)} - v$	$\theta^{(i)} - v - g_{(1)}s$
1	(− .9, .58)	(−3.7, −.30)	(−2.5, .07)
2	(−2.5, .07)	(−26.7, −.07)	(−8.9, .14)
3	(−8.9, .14)	(−3.5, .18)	(−7.9, .05)

Even in the first example of this section, the correction provides a significant improvement in the first Newton–Raphson step.

5. Discussion

The convergence of the method described in this paper ultimately depends on the accuracy of the quadratic approximation of l using its first two derivatives. Since this is the basis of the Newton–Raphson algorithm and some others for maximising the

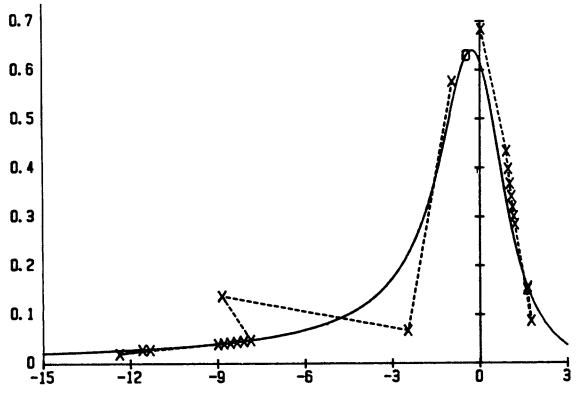

Fig. 2. Solid curve is $\{(\theta_1, \hat{\omega}(\theta_1))^t\}$, 0 gives the location of $\hat{\theta}$, and x represent points $\theta^{(i)}$ from the computation of the confidence interval for θ_1, in the example from Williams (1986)

log-likelihood, convergence properties should be very much the same. We must assume that there exists a neighbourhood of a confidence interval endpoint θ^* within which iterations like equation (9) will converge to θ^*. This neighbourhood is of undetermined size, though it certainly does not contain $\hat{\theta}$. The first step attempts to reach a point in this neighbourhood. Some prior information on the accuracy of the approximation for a particular application is usually available from the computation of the m.l.e. Numerical errors in computing derivatives, etc., will similarly be no more serious for confidence intervals than they were for the m.l.e. The only additional computation requiring a little care is the solution of the quadratic equation (8).

The important choice of a starting point $\theta^{(1)}$ for the modified Newton–Raphson iterations can be made with more sophisticated methods. As stated before, θ^* lies on the curve $\{(\beta, \hat{\omega}(\beta)^t)^t\}$. The first step away from $\hat{\theta}$ given by equation (4) is a special case of the numerical solution of the differential equation defining $\hat{\omega}(\beta)$, with β as the independent variable. Equation (4) employs one step with the elementary Euler's method (Braun, 1975, p. 136). For highly non-quadratic log-likelihoods, several steps with a smaller step size h or more accurate techniques for solving differential equations could be used to find a better starting point. We expect that this procedure would still involve less computational effort than finding the zeroes of $\hat{l}_j(\beta) - l^*$ because in such pathological cases, the maximisation over the other parameters is often difficult. Further investigation of step sizes, etc., would be needed for the extension to such a hybrid algorithm.

For even more generality, methods other than Newton–Raphson might be applied to the solution of system (2). The observed second derivatives of l can sometimes be replaced by their expectations, for instance. So long as there is an adequate quadratic approximation leading to an equation like (7), the modified step in equation (9) should still enhance convergence.

Computer-intensive methods of confidence interval construction are becoming increasingly practical these days. The bootstrap method (Efron, 1979) and its variations are well known examples. Intervals based on variance stabilising parameter transformations can also be obtained at the cost of numerically solving a system of k differential equations (Moolgavkar and Venzon, 1987b). These methods are most appropriate for parameters in non-linear models and with small sample sizes, and thus their properties can often be determined only by simulation. Efficient computation is then a high priority. Comparative studies of computation time and accurate coverage probabilities in important cases might now be of use.

A different approach for log-likelihoods with a scalar parameter is to transform the parameter in order to improve the properties of confidence intervals symmetric about the m.l.e. (Sprott, 1973; Hougaard, 1982; DiCiccio, 1984; Prentice and Mason, 1986). These transformations and confidence intervals for transformed parameters may sometimes be computed quickly and directly by evaluating certain expressions at the m.l.e. Corresponding parameterisations in the multiparameter case are more difficult to implement, however (Kass, 1984; Moolgavkar and Venzon, 1987a).

Acknowledgements

This research was supported by USPHS grant CA-39949 from the National Institutes of Health, USA. The authors wish to thank the referees for suggesting several improvements in the presentation of this paper.

Appendix

The following is a list of steps for implementing the algorithm described in Section 2 to compute the profile-likelihood-based confidence intervals for the jth parameter in the vector θ. The references to equations (3), (4) and (5) are to be understood as applying to the generalised forms, with β, $l(\theta) - l^*$ and its derivatives occupying the jth rows of the various matrices and vectors, as in step 5.

1. With $\theta = \hat{\theta}$, compute $\partial l/\partial\theta$ and $\partial^2 l/\partial\theta^2$. Store $\partial^2 l/\partial\theta^2$ as D_0.
2. Using appropriate submatrices of $\partial^2 l/\partial\theta^2$, compute $d\hat{\omega}/d\beta$ (equation (5)) and h (equation (6)).
3. Set $i = 1$ and compute $\theta^{(1)}$ (equation (4)).
4. With $\theta = \theta^{(i)}$, compute $\partial l/\partial\theta$ and $\partial^2 l/\partial\theta^2$. Store $\partial^2 l/\partial\theta^2$ as D.
5. Replace row j of $\partial^2 l/\partial\theta^2$ with $\partial l/\partial\theta$, and then replace entry j of $\partial l/\partial\theta$ with $l(\theta^{(i)}) - l^*$.
6. Compute the Newton–Raphson step v (second term, right-hand side, equation (3)).
7. Using D, $g_{(j)}$ and v, compute solution s to equation (8), choosing the root that minimises $(v + sg_{(j)})^t D_0 (v + sg_{(j)})$. If no real solution exists, invoke some method to take a conservative step, and go to step 9.
8. Compute $\theta^{(i+1)} = \theta^{(i)} - v - g_{(j)}s$.
9. If convergence is not yet attained, set i to $i + 1$, and go to step 4.
10. Return $(\theta^{(i+1)})_j$ as one endpoint of the interval. If the second endpoint remains to be found, change the sign of h and go to step 3.

References

Braun, M. (1975) *Differential Equations and Their Applications.* New York: Springer.

Breslow, N. E. and Day, N. E. (1980) *Statistical Methods in Cancer Research, Volume 1 – The Analysis of Case-Control Studies.* Lyon: International Agency for Research on Cancer.

Cox, D. R. (1970) *Analysis of Binary Data.* London: Methuen.

DiCiccio, T. J. (1984) On parameter transformations and interval estimation. *Biometrika*, **71**, 477–485.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.

Hougaard, P. (1982) Parametrizations of nonlinear models. *J. R. Statist. Soc.* B, **44**, 244–252.

Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kass, R. (1984) Canonical parameterizations and zero parameter-effects curvature. *J. R. Statist. Soc.* B, **46**, 86–92.

Moolgavkar, S. H. and Venzon, D. J. (1987a) Confidence regions in curved exponential families: application to matched case-control and survival studies with general relative risk function. *Ann. Statist.*, **15**, 346–359.

———(1987b) Confidence intervals for parameters of the proportional hazards model: a simulation study. *Scand. J. Statist.*, **14**, 43–56.

Prentice, R. L. and Mason, M. W. (1986) On the application of linear relative risk models. *Biometrics*, **42**, 109–120.

Sprott, D. A. (1973) Normal likelihoods and their relation to large sample theory of estimation. *Biometrika*, **60**, 457–466.

Storer, B. E., Wacholder, S. and Breslow, N. E. (1983) Maximum likelihood fitting of general risk models to stratified data. *Appl. Statist.*, **32**, 172–181.

Williams, D. A. (1986) Interval estimation of the median lethal dose. *Biometrics*, **42**, 641–645.