Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ESTIMATING FUNCTIONS IN FAILURE TIME DATA ANALYSIS

Ross L. Prentice and Li Hsu
Fred Hutchinson Cancer Research Center
Seattle

ABSTRACT

The relative risk model of Cox (1972) has become the standard for the regression analysis of univariate failure time data. Cox's maximum partial likelihood estimator is shown to arise from a mean parameter estimating function for the cumulative baseline hazard variate, after allowing for right censorship and upon inserting the usual baseline hazard function estimator. Mean and covariance parameter estimating functions applied to these same cumulative baseline hazard variates lead to estimation procedures for the regression analysis of multivariate failure time data. For example these estimating procedures may be used to simultaneously estimate marginal hazard ratio parameters and pairwise cross ratio parameters. Such estimation allows assessment of regression effects on marginal hazard functions while providing summary measures of the strength of dependency among pairs of failure time variates. Some additional topics in the analysis of multivariate failure time data are briefly discussed.

1 Introduction

Let $T \geq 0$ be a failure time variate. The distribution of T can be represented by the hazard (differential) function $\Lambda(dt) = E\{N(dt)|T \geq t\}$, where N is the failure time counting process corresponding to T, defined by $N(t) = 1$ if $T \leq t$ and $N(t) = 0$ otherwise, and where E denotes expectation. Note that $\Lambda(dt) = \lambda(t)dt$, where λ is the hazard function, if T is absolutely continuous. The distribution of T is also determined by the cumulative hazard function $\Lambda(t) = \int_0^t \Lambda(dt)$ or by the survivor function

$$F(t) = \mathrm{pr}(T \geq t) = \prod_{s<t} \{1 - \Lambda(ds)\}.$$

where $\prod$ denotes a product integral. Corresponding to the counting process N one can define a martingale, M, by

$$M(t) = N(t) - \Lambda(T \wedge t),$$

where '$\wedge$' denotes minimum. See Fleming and Harrington (1992) and Andersen et al (1993) for further detail and extensions.

Suppose now that an absolutely continuous failure time T is accompanied by a regression p-vector $Z = (Z_1, \ldots, Z_p)'$ and that the Cox (1972) regression model, $\Lambda(dt) = \Lambda_\circ(dt)\exp(Z'\beta)$, holds where β is a p-vector of 'relative risk' parameters and $\Lambda_\circ$ is an unspecified baseline hazard (differential) function. The regression parameter β can be estimated by the maximum partial likelihood estimator which solves

$$\sum_{k=1}^{K} Z_k \ \hat{M}_k(X_k) = 0, \tag{1}$$

based on data $(X_k, \delta_k, Z_k), k = 1, \ldots, K$, where $X_k = T_k \wedge C_k$ and $\delta_k = I[X_k = T_k]$, and C_k is a right censoring time such that T_k and C_k are independent given Z_k. Also the '$\wedge$' in (1) indicates that $\Lambda_\circ$ has been replaced by a standard baseline hazard function estimator (Breslow, 1974; Andersen and Gill, 1982)

$$\hat{\Lambda}_\circ(dt) = \sum_{k \epsilon R(t)} N(dt) / \sum_{k \epsilon R(t)} e^{Z_k' \beta}, \tag{2}$$

where $R(t) = \{k | X_k \geq t\}$ is the 'risk set' at time t. The estimating equation (1) has been justified via partial likelihood (Cox, 1975), marginal likelihood (Kalbfleisch and Prentice, 1973) and full likelihood (Johansen, 1978) arguments. The solution $\hat{\beta}$ has been shown to be semiparametric efficient (Begun et al, 1983).

Expression (1) can also be derived from the standard mean parameter estimating function

$$\sum_{k=1}^{K} D_k' \ \Sigma_k^{-1} \ (y_k - \mu_k) = 0 \tag{3}$$

for a scalar or vector response y_k having mean $\mu_k = \mu_k(\beta)$ that depends on a parameter of interest β and variance $\sum_k = \sum_k(\beta)$, and where $D_k = \partial\mu_k / \partial\beta'$. Note that (3) can be derived as the score (maximum likelihood) estimating equation for β under an exponential family model if y_k is a scalar, and as the score equation under a 'partly exponential' family more generally (Zhao, Prentice and Self, 1992). For application to uncensored univariate failure time data under the Cox model we can set $y_k = \Lambda_\circ(T_k), k = 1, \ldots, K$, in which case y_k is exponentially distributed with a mean $\exp(-Z_k'\beta)$, variance $\exp(-2Z_k'\beta)$, and $D_k = Z_k'\exp(-Z_k'\beta)$, so that (3) can be written

$$\sum_{k=1}^{K} Z_k\{1 - \Lambda_k(T_k)\} = 0.$$

This equation can also be written

$$\sum_{k=1}^{K} Z_k \ M_k(T_k) = \sum_{k=1}^{K} Z_k \int_0^{T_k} M_k(dt) = 0,$$

upon noting that $N_k(T_k) \equiv 1$ and $M_k(T_k) = 1 - \Lambda_k(T_k)$. Under independent right censorship one can integrate the martingale M_k only from zero to X_k giving the estimating function $\sum Z_k \ M_k(X_k)$. Insertion of the baseline hazard function estimator (2) then gives the estimating equation (1). One motivation for this development is to provide the basis for extension to the regression analysis of multivariate failure time data, by applying mean or mean and covariance parameter estimating equations to cumulative baseline hazard variates.

2 Estimating Functions for Marginal Hazard Ratio Parameters

While univariate failure time methods, including Kaplan-Meier curves, censored data rank tests, and Cox regression methods are well developed, multivariate failure time methods require much further development.

Suppose that there are n absolutely continuous failure time variates $(T_1, \ldots, T_n)$ and that each T_i is accompanied by a regression p-vector $Z_i, i = 1, \ldots, n$. Much of the work on the regression analysis of multivariate failure time data assumes the T_i to be independent conditional on covariates and on the value of a hypothetical frailty variate, that is usually assumed to affect the hazard function in a multiplicative manner (e.g., Andersen et al, 1993, Chapters 9 and 10; Bickel et al, 1993, Chapters 4, 6 and 7). The joint survivor function for $(T_1, \ldots, T_n)$ is obtained by integrating out the frailty variable; and dependency among the failure times are characterized by the parameters of the frailty distribution. Cox (1972) models are typically specified for the hazard functions, conditional on frailty and covariates.

Though frailty models undoubtedly have a place in multivariate failure time analysis they are limited in the flexibility with which dependencies can be modeled, since the parameters of a single frailty variate characterize the entire dependence structure for $T_1, \ldots, T_n$. More complicated frailty constructions with certain frailty variates shared by some, but not all, failure times in a correlated set may be able to partially overcome this limitation. A second problem concerns the form of the marginal distributions. Specifically, if Cox model forms are assumed for intensities given the frailties, the marginal intensities will generally no longer be of the Cox model form. Hence, in using a frailty model approach for multivariate data one may find oneself fitting marginal hazard models that differ from those that would be used if only the data on a specific margin were available. For these reasons a modeling approach that focuses on marginal survivor functions and pertinent pairwise dependency functions may be preferred.

Consider failure time variates $T_k = (T_{k1}, \ldots, T_{kn})'$ that are subject to right censorship by potential censoring variates $C_k = (C_{k1}, \ldots, C_{kn})'$ such that T_k and C_k are independent given the corresponding matrix $Z_k = (Z_{k1}, \ldots, Z_{kn})'$ of regression vectors, $k = 1, \ldots, K$. Suppose now that each T_{ki} has a marginal hazard rate function of Cox model form

$$\Lambda_{ki}(dt_{ki}) = \Lambda_{oi}(dt_{ki})\exp(Z'_{ki}\ \beta),$$

where Λ_{ki} conditions on Z_k and C_k and on the evolving failure time information, $N_{ki}(u),\ u < t_{ki}$, for the $(k,i)^{th}$ individual. Define $y_{ki} = \Lambda_{oi}(T_{ki})$, so that y_{ki} has mean $\exp(-Z'_{ki}\beta)$ and variance $\exp(-2\ Z'_{ki}\beta)$ for all (k,i). Denote by $\rho_{kij} = \rho_{kij}(\beta,\alpha)$ the correlation between y_{ki} and y_{kj}, so that the covariance $\sum_{kij} = \sum_{kij}(\beta,\alpha)$ between y_{ki} and y_{kj} is $\exp(-Z'_{ki}\beta)\exp(-Z'_{kj}\beta)\rho_{kij}$. Hence the variance matrix for y_k can be written

$$\Sigma_k = \text{diag}\{\exp(-Z'_{k1}\beta), \cdots, \exp(-Z'_{kn}\beta)\}\Omega_k\ \text{diag}\{\exp(-Z'_{k1}\beta), \cdots, \exp(-Z'_{kn}\beta)\}$$

where $\Omega_k = \Omega_k(\beta,\alpha)$ is the correlation matrix (ρ_{kij}). Also the partial derivative of $\mu_k(\beta)$ with respect to β' can be written

$$D_k = Z'_k\ \text{diag}\{\exp(-Z'_{k1}\beta), \cdots, \exp(-Z'_{kn}\beta)\},$$

so that (3) reduces to

$$\sum_{k=1}^{K} Z_k\ \Omega_k^{-1}\ M_k(T_k) = 0$$

in the absence of censorship, where $M_k(T_k) = \{M_{k1}(T_{k1}), \ldots, M_{kn}(T_{kn})\}'$. To accommodate right censorship one can replace $M_k(T_k)$ by $M_k(X_k)$ where $X_k = (X_{k1}, \ldots, X_{kn})'$. To accommodate unknown baseline hazard functions we can insert estimators (2) for each of $i = 1, \ldots, n$. It also seems natural to replace Ω_k, the correlation matrix for $M_k(T_k)$ by the correlation matrix, say $\Delta_k = \Delta_k(\beta, \alpha, C_k)$ for $M_k(X_k)$, giving the estimating equation

$$\sum_{k=1}^{K} Z_k\ \hat{\Delta}_k^{-1}\ \hat{M}_k(X_k) = 0 \tag{4}$$

for the marginal hazard ratio parameter β, where the $\wedge$ on Δ_k again denotes that the baseline hazard function estimators (2) have been inserted. Following Liang and Zeger (1986) we may consider the use of (4) with the correlation matrix replaced by a working correlation matrix, say $R_k = R_k(\beta,\alpha)$.

Provided a $K^{\frac{1}{2}}$ consistent estimator $\hat{\alpha} = \hat{\alpha}(\beta)$ is available one may then estimate β as solution to

$$\sum_{k=1}^{K} Z_k \; R_k^{-1}\{\beta, \hat{\alpha}(\beta)\} \; \hat{M}_k(X_k) = 0. \tag{5}$$

Note, however, that some judgment is required in specifying the working correlation matrix, as $\hat{\alpha}$ may not be convergent to any fixed parameter as $K \to \infty$ if the working and true matrices are too disparate (Crowder, 1995).

Wei, Lin and Weissfeld (1989) considered (5) in this multivariate failure time context with identity working correlation matrix $R_k = I_k$. Cai and Prentice (1995) considered an estimator that solves (5) with R_k a nonparametric estimate of the correlation matrix for $M_k(X_k)$, for bivariate failure time data ($n = 2$) and regression matrix Z_k that has finite support. Simulation studies conducted under the bivariate survival model of Clayton (1978) and Clayton and Cuzick (1985) indicated that the inclusion of this weight matrix did not appreciably improve the efficiency of $\hat{\beta}$ solving (5) unless the dependency between the failure times (T_{k1}, T_{k2}) was quite strong (e.g., $\rho_{k12} \geq 0.5$), and furthermore that right censorship tends to reduce any such efficiency gain. These exercises then suggest that the simple estimating equation of Wei, Lin and Weissfeld, given by

$$\sum_{k=1}^{K} Z_k \; \hat{M}_k(X_k) = 0 \tag{6}$$

will be efficient enough for marginal hazard ratio estimation in most applications. Asymptotic distribution theory showing $\hat{\beta}$ solving (5) or (6) to be consistent and asymptotically normally distributed, and including a 'sandwich' variance estimator for $\hat{\beta}$ has been presented (Wei et al, 1989; Cai and Prentice, 1995). Time-varying covariates can be accommodated by generalizing (6) to

$$\sum_{k=1}^{K} \int_0^{X_k} \; Z_k(t)\hat{M}_k(dt) = 0$$

where integration takes place componentwise for the elements of $t = (t_1, \ldots, t_n)'$.

3 Estimating Functions for Hazard Ratio and Pairwise Cross Ratio Parameters

In many multivariate failure time applications it will be important to not only estimate marginal hazard rates, but also to develop summary measures of the strength of dependence among pairs of failure times. For example, dependency measures may be of primary interest in some contexts, for example in studies of disease occurrence among family members in genetic epidemiology.

The 'cross-ratio' function (e.g., Oakes, 1982, 1986, 1989)

$$CR_{ij}(t_k, t_j; Z) = \frac{F_{ij}(dt_i, dt_j; Z)\ F_{ij}(t_i, t_j; Z)}{F_{ij}(dt_i, t_j; Z)\ F_{ij}(t_i, dt_j; Z)}$$

provides a useful characterization of the relationship between failure time variates T_i and T_j. Note that CR_{ij} can also be expressed as

$$\begin{aligned} CR_{ij}(t_i, t_j; Z) &= \lambda_i(t_i|T_j = t_j; Z)/\lambda_i(t_i|T_j \geq t_j; Z) \\ &= \lambda_j(t_j|T_i = t_i; Z)/\lambda_j(t_j|T_i \geq t_i; Z), \end{aligned}$$

which has a very natural interpretation in epidemiologic and other contexts. The Clayton model (Clayton, 1978; Clayton and Cuzick, 1985) supposes that the cross ratio is a constant

$$CR_{ij}(t_i, t_j; Z) = 1 + \theta_{ij}(Z)$$

for all (t_i, t_j), in which case $\theta_{ij}(Z) \geq -0.5$ provides a summary measure of the strength of dependence between T_i and T_j, given Z, with positive and negative dependencies given by $\theta_{ij}(Z) > 0$ and $\theta_{ij}(Z) < 0$, respectively.

Now consider joint estimating functions for the marginal hazard ratio parameter β and for an additional parameter α that characterizes the correlations among cumulative hazard variates $y_{ki} = \Lambda_{oi}(T_{ki}), i = 1, \ldots, n, k = 1, \ldots, K$. Let $\sigma_k(\beta, \alpha) = (\sum_{k11}, \sum_{k12}, \ldots, \sum_{k1n}, \sum_{k22}, \ldots, \sum_{k2n}, \ldots)$ denote the variance matrix for $y_k = (y_{k1}, \ldots, y_{kn})'$ in vector form. Under a quadratic exponential model for $y_k, k = 1, \ldots, K$ the score equations for (β, α) can be written (e.g., Prentice and Zhao, 1991) as

$$\sum_{k=1}^{K} D_k'\ \Delta_k^{-1}\ f_k = 0$$

where

$$D_k = \begin{pmatrix} \partial\mu_k/\partial\beta' & 0 \\ \partial\sigma_k/\partial\beta' & \partial\sigma_k/\partial\alpha' \end{pmatrix}, \ \Delta_k = \begin{pmatrix} \sum_k & \text{cov}(y_k, s_k) \\ \text{cov}(s_k, y_k) & \text{var } s_k \end{pmatrix}, \ f_k = \begin{pmatrix} y_k - \mu_k \\ s_k - \sigma_k \end{pmatrix}$$

and where $s_k = (s_{k11}, \ldots, s_{k1n}, s_{k22}, \ldots, s_{k2n}, \ldots)$, with $s_{kij} = (y_{ki} - \mu_{ki})(y_{kj} - \mu_{kj})$, is an empirical covariance vector. Note that $E(s_k) = \sigma_k$. These mean and covariance estimating equations are attractive in that they arise as maximum likelihood equations under a rich quadratic exponential class for y_k. A drawback, however, is that misspecification of the covariance model $\sigma_k(\beta, \alpha)$ can bias the estimator of the hazard ratio parameter β. This can be remedied by replacing $\partial\sigma_k/\partial\beta'$ and $\text{cov}(y_k, s_k)$ by zero matrices giving the simplified estimating equations

$$\sum_{k=1}^{K} (\partial\mu_k'/\partial\beta)\ \Sigma_k^{-1}\ (y_k - \mu_k) = 0,\ \sum_{k=1}^{K} (\partial\sigma_k'/\partial\alpha)(\text{var } s_k)^{-1}(s_k - \sigma_k) = 0. \quad (7)$$

Under Cox-model marginal hazard functions and no censorship the first of these equations can be written (Prentice and Hsu, 1996), as before, as

$$\sum_{k=1}^{K} Z_k\ \Omega_k^{-1}\ M_k(T_k) = 0,$$

while the second equation similarly simplifies to

$$\sum_{k=1}^{K} E_k \Phi_k^{-1}\ L_k(T_k) = 0$$

where $E_k = \partial\rho_k'/\partial\alpha, \Phi_k$ is the covariance matrix for

$$\{\Lambda_{k1}(T_{k1})\Lambda_{k2}(T_{k2}), \Lambda_{k1}(T_{k1})\Lambda_{k3}(T_{k3}), \ldots\}$$

and

$$L_k(T_k) = \{L_{k12}(T_{k1}, T_{k2}), L_{k13}(T_{k1}, T_{k3}), \ldots\}$$

with

$$L_{kij}(T_{ki}, T_{kj}) = M_{ki}(T_{ki})\ M_{kj}(T_{kj}) - \rho_{kij}.$$

Note that the expectation of both estimating functions is a zero vector even if Ω_k and Φ_k are misspecified.

We can adapt these estimating functions to independent right censorship by again inserting baseline hazard function estimators (2) for $i = 1, \ldots, n$ and by replacing T_k by X_k. One may also replace the correlation matrices Ω_k and Φ_k by working covariance matrices, say R_{k1} and R_{k2}, for M_k and L_k respectively, giving the estimating equations

$$\sum_{k=1}^{K} Z_k\ \hat{R}_{k1}^{-1}\ \hat{M}_k(X_k) = 0, \sum_{k=1}^{K} E_k\ \hat{R}_{k2}^{-1}\ \hat{L}_k(X_k) = 0. \quad (8)$$

This notation conceals one important point. For $L_k(X_k)$ to have mean zero under right censorship one must redefine

$$L_{kij}(X_{ki}, X_{kj}) = M_{ki}(X_{ki})\ M_{kj}(X_{kj}) - A_{kij}(X_{ki}, X_{kj}) \quad (9)$$

where

$$A_{kij}(X_{ki}, X_{kj}) = \int_0^{X_{ki}} \int_0^{X_{kj}} A_{kij}(dt_i, dt_j)$$

and

$$A_{kij}(dt_i, dt_j) = E\{M_{ki}(dt_i), M_{kj}(dt_j) | T_{ki} \geq t_i, T_{kj} \geq t_j, Z_k\}.$$

In fact the 'covariance rate function' A_{kij} in conjunction with Λ_{ki} and Λ_{kj} completely determines the distribution of T_{ki} and T_{kj}, given Z_k (Prentice and Cai, 1992). Hence to apply (8) with right censoring one must specify the pairwise survivor functions $F_{ij}(t_i, t_j; Z_k)$. Such further assumption seems natural as the cumulative hazard variate correlations are not even identifiable if censorship restricts the support of (X_{ki}, X_{kj}).

One way to implement (8) is to impose constant cross ratio models

$$CR_{kij}(t_i, t_j; Z_k) = 1 + \theta_{kij}(\alpha)$$

for each pair of failure time variates $(T_{ki}, T_{kj}); i, j = 1, \ldots, n$, even though the existence of an overall survivor function $F(t_1, \ldots, t_n; Z_k)$ having these pairwise marginals is yet to be demonstrated. These constant cross ratio assumptions give

$$A_{kij}(dt_i, dt_j) = A_\circ\{\Lambda_{ki}(t_i), \Lambda_{kj}(t_j); \theta_{kij}\}\Lambda_{ki}(dt_i)\Lambda_{kj}(dt_j) \tag{10}$$

where

$$A_\circ(v_1, v_2; \theta) = (\theta + 1)\ e^{v_1\theta}\ e^{v_2\theta}\ (e^{v_1\theta} + e^{v_2\theta} - 1)^{-2} - (e^{v_1\theta} + e^{v_2\theta} - 1)^{-1}.$$

The cumulative hazard correlation ρ_{kij} is linked in a one–to–one fashion to the cross ratio parameter θ_{kij} via

$$\rho_{kij} = \int_0^\infty \int_0^\infty (e^{v_1\theta_{kij}} + e^{v_2\theta_{kij}} - 1)^{-\theta_{kij}^{-1}} - 1,$$

thereby determining E_k in (8). Independence working matrices $\hat{R}_{k1} = I_n$ and $\hat{R}_{k2} = I_{n(n-1)/2}$ can be expected to yield estimators of β and α of acceptable efficiency in most applications. Note that the cross ratio θ_{kij} can be modeled in various ways. For example, one could set $\theta_{kij} = \alpha_{ij}$ for all (i, j), could restrict some elements of α_{ij} to be common, or could allow θ_{kij} to depend on Z_k. $\hat{L}_k(X_k)$ in (8) arises by inserting baseline estimators (2) into (9). In doing so, simulation studies suggest the use of Kaplan Meier-type estimators

$$\hat{\Lambda}_{ki}(t_i) = \sum_{\{\ell | X_{\ell i} \leq t_i, N_{\ell i}(X_{\ell i}) = 1\}} \exp\{-Z'_{\ell i}\beta\} \log\{1 - \hat{\Lambda}_{\ell i}(d\ X_{\ell i})\}$$

in (10), where $\hat{\Lambda}_{\ell i}(dt_i) = \exp\{Z'_{\ell i}\beta\}\hat{\Lambda}_{oi}(dt_i)$.

See Prentice and Hsu (1996) for simulations and illustrations of the use of estimating equations

$$\sum_{k=1}^{K} Z_k\ \hat{M}_k(X_k) = 0, \sum_{k=1}^{K} E_k\ \hat{L}_k(X_k) = 0 \tag{11}$$

for hazard ratio parameter (β) and pairwise cross ratio parameter (α) estimation. The estimator of β solving (11) is that of Wei, Lin and Weissfeld (1989), while the estimator of α has been shown in simulation studies to have comparable efficiency to the generalized maximum likelihood estimator of Nielson et al (1992) in the non-regression special case (Hsu and Prentice, 1996). Asymptotic distribution theory, including a consistent variance estimator is available for solutions to (11) and to the more general estimating equations (8) (Prentice and Hsu, 1996).

4 Discussion

Mean parameter estimating functions can be adapted to allow for right censorship, yielding the maximum partial likelihood estimator of the hazard ratio parameter in Cox's failure time regression model, and yielding the Wei, Lin and Weissfeld (1989) estimator of marginal hazard ratio parameters under an independence working model for a multivariate failure time response. Mean and covariance model estimating functions can extend the regression analysis of multivariate failure time data to the joint estimation of marginal hazard ratio parameters and pairwise cross ratio parameters. These estimating equations (11) involve straightforward computations and the estimated parameters have a ready interpretation.

The principal limitation of the estimating equations (11) relates to the constant cross-ratio assumptions. Under departures from a constant cross ratio the estimates $\theta_{kij}(\hat{\alpha})$ presumably have an average cross ratio interpretation, with averaging over the density of $X_{kij}, k = 1, \ldots, K$. As such the interpretation of $\theta_{kij}(\hat{\alpha})$ will unfortunately depend on the censoring distribution, just as the interpretation to the solution to (6) will depend on the censorship under departure from the Cox model. The possibility of estimating average cross ratios, with averaging over the density of the underlying failure times, rather than the observed failure or censoring times, is currently being pursued by graduate student Juan Juan Fan in conjunction with the authors.

Acknowledgement

This work was supported by NIH grant CA-53996.

References

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* 10, 1100-1120.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Begun, J.M., Hall, W.J., Huang, W.M. and Wellner, J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* 11, 432-452.

Bickel, P.J., Klassen, C.A., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore, Maryland.

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics* 30, 89-99.

Cai, J. and Prentice, R.L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* 82, 151-164.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65, 144-151.

Clayton, D.G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J.R. Statist. Soc. A* 148, 82-117.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J.R. Statist. Soc. B* 187-220.

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Hsu, L. and Prentice, R.L. On assessing the strength of dependency between failure time variates. In press, *Biometrika*, 1996.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* 5, 195-199.

Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* 60, 267-278.

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

Liang, K.Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

Nielsen, G.G., Gill, R.D., Andersen, P.K. and Sorensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* 19, 25-43.

Oakes, D. (1982). A model for association in bivariate survival data. *J.R. Statist. Soc. B* 11, 111-122.

Oakes, D. (1986). Semi-parametric inference in a model for association in bivariate survival data. *Biometrika* 73, 353-361.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* 84, 487-493.

Prentice, R.L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* 79, 495-512.

Prentice, R.L. and Zhao, L.P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* 47, 825-839.

Prentice, R.L. and Hsu, L. Regression on hazard ratios and cross ratios in multivariate failure time analysis. In press, *Biometrika*, 1996.

Wei, L.J., Lin, P.Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *J. Amer. Statist. Assoc.* 84, 1065-1073.

Zhao, L.P., Prentice, R.L. and Self, S.G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *J.R. Statist. Soc. B* 54, 805-811.