Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions

L. J. Wei; D. Y. Lin; L. Weissfeld

Journal of the American Statistical Association, Vol. 84, No. 408 (Dec., 1989), 1065-1073.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198912%2984%3A408%3C1065%3ARAOMIF%3E2.0.CO%3B2-Q

Journal of the American Statistical Association is currently published by American Statistical Association.

Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions

L. J. WEI, D. Y. LIN, and L. WEISSFELD*

Many survival studies record the times to two or more distinct failures on each subject. The failures may be events of different natures or may be repetitions of the same kind of event. In this article, we consider the regression analysis of such multivariate failure time observations. Each marginal distribution of the failure times is formulated by a Cox proportional hazards model. No specific structure of dependence among the distinct failure times on each subject is imposed. The regression parameters in the Cox models are estimated by maximizing the failure-specific partial likelihoods. The resulting estimators are shown to be asymptotically jointly normal with a covariance matrix that can be consistently estimated. Simultaneous inferential procedures are then proposed. Extensive Monte Carlo studies indicate that the normal approximation is adequate for practical use. The new methods allow time-dependent covariates, missing observations, and arbitrary patterns of censorship. They are illustrated with two real-life examples. For recurrent failure time data, various regression methods have been proposed in the literature. These methods, however, generally assume stringent structures of dependence among the recurrences of each subject. Moreover, as shown in the present article, they are rather sensitive to model misspecification.

KEY WORDS: Cox model; Martingale; Partial likelihood; Proportional hazards; Simultaneous inference.

1. INTRODUCTION

The regression analysis of survival data when there is a single and possibly censored event time for each study subject has been extensively investigated [see Cox and Oakes (1984) and Kalbfleisch and Prentice (1980)]. Many studies, however, involve the recording of times to two or more distinct events or failures on each subject. The failures may be repetitions of the same kind of event or may be events of different natures. For example, in a randomized clinical trial to evaluate the effectiveness of the drug ribavirin, patients with acquired immune deficiency syndrome (AIDS) were randomly assigned to one of three groups: placebo, low-dose ribavirin, and high-dose ribavirin. One of the main interests in the study was to investigate the antiretroviral capability of ribavirin over time (see Makuch and Parks 1988). Blood samples for each patient were collected at weeks 4, 8, and 12. For each serum sample, measurements of p24 antigen levels, which are important markers of HIV-1 infection, were repeatedly taken for a period of four weeks. The "viral load" in each serum sample was evaluated by measuring the number of days when virus positivity was detected, that is, when the p24 level was greater than 100 picograms per milliliter. Therefore, potentially each patient in the study should have three such event times. Some observations were missing, however, because patients did not make the scheduled visits or because serum specimens were inadequate for laboratorial analysis. In addition, censored observations occurred when the culture required a longer period of time to register as virus positive than was achievable in the laboratory, or when the serum sample was contaminated before positivity was detected. These incomplete multivariate lymphocyte culture data are presented in Table 1. Based on these virological data, one would like to know, for example, whether the drug ribavirin effectively prolonged the time to virus positivity and how the drug effects changed over time.

Another interesting example can be found in a bladder cancer study (see Byar 1980) conducted by the Veterans Administration Cooperative Urological Research Group. In this study, all patients had superficial bladder tumors when they entered the trial. These tumors were removed transurethrally and patients were randomly assigned to one of three treatments: placebo, thiotepa, and pyridoxine. Many patients had multiple recurrences of tumors during the study, and new tumors were removed at each visit. Table 2 presents the recurrence times of tumors from Group 1 (placebo) and Group 2 (thiotepa). Because of the sparseness of the data beyond the fourth recurrence, only the first four recurrence times are reported. Here, each recurrence time of a patient was measured from the beginning of his treatment. As indicated in Byar (1980), one of the analyses to evaluate the effectiveness of thiotepa should be based on the tumor recurrence times from patients in these two groups.

Several regression methods have been proposed in the literature to deal with situations where individuals experience repeated failures such as multiple tumor recurrences. These methods impose specific structures of dependence among the recurrences on each subject and can be thought of as generalizations of survival data techniques in which the hazard function modeling is continued beyond a subject's first failure to the second and subsequent fail-

* L. J. Wei is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. D. Y. Lin is Postdoctoral Research Fellow, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. L. Weissfeld is Assistant Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109. The authors thank Robert W. Makuch for providing the AIDS data. They are also grateful to an associate editor, two referees, John D. Kalbfleisch, and Robert Wolfe for their insightful comments. This research was supported in part by National Institutes of Health Grant RO1-CA45544.

Journal of the American Statistical Association
December 1989, Vol. 84, No. 408, Theory and Methods

Table 1. Days to Virus Positivity of Serum Samples From AIDS Patients

Treatment	Month 1	Month 2	Month 3
1	9	6	7
1	4	5	10
1	6	7	6
1	10	—	21*
1	15	8	—
1	3	—	6
1	4	7	3
1	9	12	12
1	9	19	19*
1	6	5	6
1	9	—	18
1	9	20*	17*
2	6	4	5
2	16	17	21*
2	31	19*	21*
2	27*	19*	—
2	7	16	23*
2	28*	7	19*
2	28*	3	16
2	15	12	16
2	18	21*	22
2	8	4	7
2	4	21*	7
3	21	9	8*
3	13	7	21*
3	16	6	20
3	3	8	6
3	21	—	25*
3	7	19	3
3	11	13	21*
3	27*	18*	9
3	14	14	6
3	8	11	15
3	8	4	7
3	8	3	9
3	19*	10	17*

NOTE: *indicates censored; — indicates missing. Treatment: 1, placebo; 2, low-dose ribavirin; 3, high-dose ribavirin.

ures. For example, Lawless (1987) presented a class of parametric and semiparametric procedures based on nonhomogeneous Poisson process models with proportional intensity assumptions. The counting process formulation of Andersen and Gill (1982) can be regarded as a special case with the Cox proportional intensity model. The approach of Prentice, Williams, and Peterson (1981) differs from that of Andersen and Gill (1982) in two aspects: (a) the risk sets for the $(k + 1)$th recurrences are restricted to the individuals who have experienced the first k recurrences; and (b) the underlying intensity functions and regression parameters are allowed to vary among distinct recurrences. The method of Gail, Santner, and Brown (1980) is a two-sample special case of Prentice et al. (1981).

In this article, semiparametric methods are proposed to analyze the general multivariate failure time data. Specifically, we model the marginal distribution of each failure time variable with a Cox proportional hazards model. No particular structure of dependence among distinct failure times on each subject is imposed. The regression parameters are estimated by maximizing the failure-specific partial likelihoods. The resulting estimators across all types of failures are shown to be asymptotically jointly normal with a covariance matrix that can be easily estimated from the data. This permits a global analysis of, for example, the effects from the drug ribavirin over the three time points based on the virological data from the three dose groups of AIDS patients. The new procedures are presented in the next section and are illustrated in Section 3 with the aforementioned two real-life examples.

2. MODELING MARGINAL DISTRIBUTIONS OF MULTIVARIATE FAILURE TIME WITH PROPORTIONAL HAZARDS MODELS

For the kth type of failure ($k = 1, \ldots, K$), let X_{ki} be the failure time of the ith subject ($i = 1, \ldots, n$). Either deliberately or accidentally, however, for $\tilde{X}_{ki}$ one observes a bivariate vector (X_{ki}, Δ_{ki}), where $X_{ki} = \min(\tilde{X}_{ki}, C_{ki})$, C_{ki} is the censoring time, and $\Delta_{ki} = 1$ if $X_{ki} = \tilde{X}_{ki}$ and 0 otherwise. If $\tilde{X}_{ki}$ is missing, let C_{ki} be 0. This implies that $X_{kk} = 0$ and $\Delta_{ki} = 0$, since $\tilde{X}_{ki}$ is positive. Now, let $Z_{ki}(t)$, $= (Z_{1ki}(t), \ldots, Z_{pki}(t))'$ denote a $p \times 1$ vector of covariates for the ith subject at time $t \geq 0$ with respect to the kth type of failure. Conditional on Z_{ki}, the failure vector $\tilde{X}_i = (\tilde{X}_{1i}, \ldots, \tilde{X}_{Ki})'$ and the censoring vector $C_i = (C_{1i}, \ldots, C_{Ki})'$ ($i = 1, \ldots, n$) are assumed to be independent. Furthermore, we assume that $(X_i, \Delta_i, Z_i(\cdot))$ ($i = 1, \ldots, n$), where $Z_i = (Z_{1i}', \ldots, Z_{Ki}')'$, are iid random vectors with bounded covariates $Z_i(\cdot)$.

For the kth type of failure of the ith subject, the hazard function $\lambda_{ki}(t)$ is assumed to take the form

$$\lambda_{ki}(t) = \lambda_{k0}(t)\exp\{\beta_k' Z_{ki}(t)\}, \qquad t \geq 0, \tag{2.1}$$

where $\lambda_{k0}(t)$ is an unspecified baseline hazard function and $\beta_k = (\beta_{1k}, \ldots, \beta_{pk})'$ is the failure-specific regression parameter. Now, let $\mathcal{R}_k(t) = \{l : X_{kl} \geq t\}$, that is, the set of subjects at risk just prior to time t with respect to the kth type of failure. Then the kth failure-specific partial likelihood (Cox 1972; Cox 1975) is

$$L_k(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\{\beta' Z_{ki}(X_{ki})\}}{\sum_{l \in \mathcal{R}_k(X_{ki})} \exp\{\beta' Z_{kl}(X_{ki})\}} \right]^{\Delta_{ki}}. \tag{2.2}$$

The maximum partial likelihood estimator $\hat{\beta}_k$ for β_k is defined as the solution to the likelihood equation $\partial \log L_k(\beta)/\partial\beta = 0$. The estimator $\hat{\beta}_k$ is consistent for β_k if Model (2.1) is correctly specified.

The estimators $\hat{\beta}_k$'s are generally correlated. As shown in the Appendix, for large n, $(\hat{\beta}_1', \ldots, \hat{\beta}_K')'$ is approximately normal with mean $(\beta_1', \ldots, \beta_K')'$ and covariance matrix Q, say. In the Appendix, we alo give an estimator $\hat{Q}$ of Q in (A.3). This provides a basis for simultaneous inferences about the β_k's. For instance, suppose that we are interested in the effects $\eta_k = \beta_{1k}$ ($k = 1, \ldots, K$) from a particular kind of covariates on the K event times. Let the estimator of the covariance matrix of $(\hat{\eta}_1, \ldots, \hat{\eta}_K)'$, where $\hat{\eta}_k = \hat{\beta}_{1k}$, be denoted by $\hat{\Psi}$, which can be obtained from $\hat{Q}$. Then the quadratic form $W = (\hat{\eta}_1, \ldots, \hat{\eta}_K)\,\hat{\Psi}^{-1}(\hat{\eta}_1, \ldots, \hat{\eta}_K)'$ can be used to test jointly the null hypotheses $H_k : \eta_k = 0$ ($k = 1, \ldots, K$). Now, suppose that $\eta_1 = \cdots = \eta_K = \eta$, then it is natural to estimate η by a linear combination of the $\hat{\eta}_k$'s, that is, $\sum_{k=1}^{K} c_k\hat{\eta}_k$ with $\sum_{k=1}^{K} c_k = 1$. The estimator $\hat{\eta}$ with weight

Table 2. Tumor Recurrence Data for Patients With Bladder Cancer

Treatment group	Follow-up time	Initial number	Initial size	Recurrence time 1	2	3	4	Treatment group	Follow-up time	Initial number	Initial size	Recurrence time 1	2	3	4
1	0	1	1					1	53	3	1	3	15	46	51
1	1	1	3					1	59	1	1				
1	4	2	1					1	61	3	2	2	15	24	30
1	7	1	1					1	64	1	3	5	14	19	27
1	10	5	1					1	64	2	3	2	8	12	13
1	10	4	1	6				2	1	1	3				
1	14	1	1					2	1	1	1				
1	18	1	1					2	5	8	1	5			
1	18	1	3	5				2	9	1	2				
1	18	1	1	12	16			2	10	1	1				
1	23	3	3					2	13	1	1				
1	23	1	3	10	15			2	14	2	6	3			
1	23	1	1	3	16	23		2	17	5	3	1	3	5	7
1	23	3	1	3	9	21		2	18	5	1				
1	24	2	3	7	10	16	24	2	18	1	3	17			
1	25	1	1	3	15	25		2	19	5	1	2			
1	26	1	2					2	21	1	1	17	19		
1	26	8	1	1				2	22	1	1				
1	26	1	4	2	26			2	25	1	3				
1	28	1	2	25				2	25	1	5				
1	29	1	4					2	25	1	1				
1	29	1	2					2	26	1	1	6	12	13	
1	29	4	1					2	27	1	1	6			
1	30	1	6	28	30			2	29	2	1	2			
1	30	1	5	2	17	22		2	36	8	3	26	35		
1	30	2	1	3	6	8	12	2	38	1	1				
1	31	1	3	12	15	24		2	39	1	1	22	23	27	32
1	32	1	2					2	39	6	1	4	16	23	27
1	34	2	1					2	40	3	1	24	26	29	40
1	36	2	1					2	41	3	2				
1	36	3	1	29				2	41	1	1				
1	37	1	2					2	43	1	1	1	27		
1	40	4	1	9	17	22	24	2	44	1	1				
1	40	5	1	16	19	23	29	2	44	6	1	2	20	23	27
1	41	1	2					2	45	1	2				
1	43	1	1	3				2	46	1	4	2			
1	43	2	6	6				2	46	1	4				
1	44	2	1	3	6	9		2	49	3	3				
1	45	1	1	9	11	20	26	2	50	1	1				
1	48	1	1	18				2	50	4	1	4	24	47	
1	49	1	3					2	54	3	4				
1	51	3	1	35				2	54	2	1	38			
1	53	1	7	17				2	59	1	3				

NOTE: Treatment group: 1, placebo; 2, thiotepa. Follow-up time and recurrence time are measured in months. Initial size is measured in centimeters. Initial number of 8 denotes eight or more initial tumors.

Source: Andrews and Herzberg (1985, pp. 254–259).

$c = (c_1, \ldots, c_K)' = (e'\hat{\Psi}^{-1}e)^{-1}\hat{\Psi}^{-1}e$, where $e = (1, \ldots, 1)'$, has the smallest asymptotic variance among all of the linear estimators (see Wei and Johnson 1985). Notice that, even if the η_k's are unequal, in practice one may still combine the $\hat{\eta}_k$'s to draw a conclusion about the "average effect" of the covariates provided that there are no qualitative differences among the η_k's.

The longitudinal failure time data, such as those described in Section 1, provide us with the opportunity to study the changes of the effects η_k's over time. If we do not prespecify any relationship among the η_k's, we are faced with a multiple inference problem. For simplicity, let us assume that $\eta_k \leq 0$ for all k. If all of the null hypotheses H_k's are true, the standardized estimator $(\tilde{\eta}_1, \ldots, \tilde{\eta}_K)'$, where $\tilde{\eta}_k = \hat{\eta}_k/\hat{\psi}_{kk}^{1/2}$, is approximately normal with mean 0 and covariance matrix $\tilde{\Psi} = \{\tilde{\psi}_{kl}\}$, where $\tilde{\psi}_{kl} = \hat{\psi}_{kl}/(\hat{\psi}_{kk}\hat{\psi}_{ll})^{1/2}$ and $\hat{\psi}_{kl}$ is the (k, l)th element of $\hat{\Psi}$. A conventional multiple testing procedure rejects H_k if $\tilde{\eta}_k < d$, where d is the largest value such that $\Pr(\tilde{\eta}_k \geq d, k = 1, \ldots, K \mid H_1, \ldots, H_K) \geq 1 - \alpha$ and α is a prespecified level of significance. The sequential multiple test procedures studied by Marcus, Peritz, and Gabriel (1976), Holm (1979), and Wei and Stram (1988), however, can be applied to the present case to obtain more powerful tests. Now, let $\tilde{\eta}_k^*$ be the kth smallest observed value of the $\tilde{\eta}_k$'s, and let $\tilde{\Psi}^*$ be the corresponding covariance matrix obtained by rearranging the rows and columns of $\tilde{\Psi}$. In addition, let $H_k^* : \eta_k^* = 0$ be the ordered hypotheses from the H_k's according to the order of $\tilde{\eta}_1^*, \ldots, \tilde{\eta}_K^*$,. Furthermore, let $(V_1, \ldots, V_K)'$ be a multivariate normal vector with mean 0 and covariance matrix $\tilde{\Psi}^*$. Starting with the hypothesis H_1^*, reject H_k^* $(k = 1, \ldots, K)$ if $\Pr(\min_{k \leq j \leq K} V_j \leq \tilde{\eta}_k^*) \leq \alpha$, provided that $H_1^*, \ldots, H_{k-1}^*$ have been tested and rejected. It can be shown that the Type I error probability of this procedure is α asymptotically for any combination of the true H_k's. The procedures for testing

two-sided alternatives can be developed in similar fashions.

Before applying the foregoing procedures to the real-life examples, we need to check whether the normal approximation to the distribution of $(\hat{\beta}_1', \ldots, \hat{\beta}_K')'$ is accurate enough for practical use. To this end, extensive Monte Carlo studies were carried out. The results indicated that the approximation was fairly accurate for moderate-sized samples. For example, in one of the studies, we considered a two-sample problem with two time points; that is, $K = 2$. The joint distribution for the two event times $\tilde{X}_1$ and $\tilde{X}_2$ was from a family of bivariate exponential distributions studied by Gumbel (1960). The distribution functions take the form $F(u, v) = F_1(u)F_2(v)[1 + \theta\{1 - F_1(u)\}\{1 - F_2(v)\}]$, where $-1 \le \theta \le 1$. The parameter θ measures the degree of dependence between the two event times in that the correlation between $\tilde{X}_1$ and $\tilde{X}_2$ equals $\theta/4$. The two marginal distribution functions $F_1(\cdot)$ and $F_2(\cdot)$ were the univariate exponentials with hazard rates $\exp(\beta_1 Z)$ and $\exp(\beta_2 Z)$, respectively, where Z was a (0, 1)-indicator. For simplicity, we let $\beta_1 = \beta_2 = \beta$. We were interested in checking whether the confidence interval of β based on the normal approximation to the distribution of the combined estimator $\hat{\beta}$ would be accurate enough for practical use. Various values of n, θ, and β were considered. Table 3 presents some typical results of these simulations. As shown in the table, the empirical coverage probabilities of the intervals are very close to the nominal confidence levels for $n \ge 50$ with various censoring patterns. In fact, the discrepancies are not alarming even for smaller samples.

For analyzing recurrent failure time data, our method provides an appealing alternative to those proposed by Prentice et al. (1981). In their approach, the hazard function at time t is defined as the instantaneous rate of recurrence at time t given the covariates and recurrence history up to that time. Their models (2) and (3) specify the underlying hazard functions as functions of the time from the beginning of the study and from the subject's immediately preceding recurrence, respectively. Stringent assumptions on the within-subject dependence are implicitly made in these two models. For example, in their model (2) with time-independent covariates, the hazard function for the $(k + 1)$th recurrence of a subject is assumed to be independent of the subject's previous k recurrence times. From another set of extensive Monte Carlo studies, we found that the procedures of Prentice et al. (1981) were rather sensitive to model assumptions. For example, in one of our studies, two gap times U and V between distinct recurrences were generated from the bivariate exponential distributions of Gumbel (1960) discussed previously. The maximum correlation coefficient between U and V from the Gumbel distributions is only .25. To study cases with higher correlation coefficients, the gap times were also generated from the bivariate exponentials studied by Lawrance and Lewis (1981). Here, for example, to generate U and V with .5 correlation, we let $U = E_1$ and $V = .5E_1 + \pi E_2$, where E_1 and E_2 were independent exponential variables with hazard rates $\exp(\beta_1 Z)$ and $\exp(\beta_2 Z)$, respectively, and π was an independent Bernoulli variable with .5 success probability. The two marginal recurrence times $\tilde{X}_1$ and $\tilde{X}_2$ were U and $U + V$, respectively. The corresponding observed quantities were $X_1 = \min\{\tilde{X}_1, C\}$ and $X_2 = \min\{\tilde{X}_2, C\}$, where C was a censoring variable. Notice that, with respect to the covariate Z, the hazard function for $\tilde{X}_2$ was no longer in the proportional hazards form unless $\beta_1 = \beta_2 = 0$. That is, if $\beta_1 \neq 0$ or $\beta_2 \neq 0$, the proportional hazards assumption for our method and the Prentice et al. model (2) was violated. It was still meaningful, however, to compare the size of the test for testing $\beta_1 = \beta_2 = 0$ based on our combined estimator $\hat{\beta}$ with those based on model (2) and model (3) of Prentice et al. (1981). The results from these studies are summarized in Table 4. Our test always maintains its size near the nominal level. The size of the test based on model (2), however, significantly exceeds the nominal level, indicating its sensitivity to model misspecification. The test based on model (3) seems valid when there is no dependence between the two gap times, but its size increases with the correlation coefficient between U and V.

3. EXAMPLES

In this section, the proposed procedures are illustrated with the two real-life examples discussed in Section 1. The first one is the study of ribavirin for AIDS patients. Here, $\tilde{X}_{ki}$ is the number of days to virus positivity in the kth serum sample of the ith patient ($k = 1, 2, 3$; $i = 1, \ldots, 36$). Let $Z_{ki} = (Z_{1ki}, Z_{2ki})'$, where $Z_{1ki} = 1$ if the ith patient was in the low-dose group, and 0 otherwise; and $Z_{2ki} = 1$

Table 3. Empirical Coverage Probabilities of the Confidence Intervals for the Common Regression Parameter $\beta(= .5)$ in the Bivariate Exponential Model With Marginal Hazards $\lambda_1(t) = \lambda_2(t) = \exp(\beta Z)$

		No censoring: Confidence level			26% Censoring[a]: Confidence level			51% Censoring[b]: Confidence level		
θ	n	90%	95%	99%	90%	95%	99%	90%	95%	99%
0	25	.881	.935	.974	.875	.925	.976	.870	.932	.979
0	50	.893	.943	.985	.892	.935	.986	.880	.940	.984
0	100	.902	.949	.988	.902	.951	.991	.896	.948	.990
1	25	.892	.939	.986	.880	.937	.985	.890	.939	.979
1	50	.908	.948	.987	.897	.941	.990	.895	.947	.982
1	100	.895	.950	.990	.904	.947	.991	.908	.944	.985

NOTE: The covariate Z is dichotomous with $\Pr(Z = 0) = \Pr(Z = 1) = .5$. Uniform random numbers are generated through an algorithm provided by Press, Flannery, Teukolsky, and Vetterling (1986, pp. 196–197). Each entry is based on 1,000 replications.

[a] Two censoring variables, both distributed as uniform(0, 3.0), are imposed independently on the two types of failures.

[b] Two censoring variables, both distributed as uniform(0, 1.2), are imposed independently on the two types of failures.

Table 4. Empirical Sizes of the Wald Tests for the Covariate Effects With Bivariate Exponential Gap Times: (i) New Method; (ii) Prentice et al. Model (2); (iii) Prentice et al. Model (3)

			No censoring: Nominal level			Uniform(0, 3) censoring[b]: Nominal level		
ρ[a]	n		.10	.05	.01	.10	.05	.01
0	50	(i)	.113	.058	.015	.110	.050	.012
		(ii)	.213	.139	.047	.147	.077	.019
		(iii)	.088	.041	.012	.096	.048	.011
0	100	(i)	.097	.047	.012	.093	.050	.011
		(ii)	.188	.122	.046	.136	.074	.025
		(iii)	.095	.043	.010	.086	.039	.011
.25	50	(i)	.100	.059	.016	.104	.056	.013
		(ii)	.208	.131	.050	.157	.079	.027
		(iii)	.138	.077	.024	.105	.060	.013
.25	100	(i)	.104	.059	.013	.107	.055	.015
		(ii)	.228	.155	.056	.161	.084	.026
		(iii)	.142	.087	.025	.117	.054	.015
.50	50	(i)	.110	.053	.015	.105	.052	.009
		(ii)	.229	.142	.051	.170	.101	.027
		(iii)	.165	.099	.039	.141	.072	.018
.50	100	(i)	.105	.052	.010	.110	.051	.009
		(ii)	.210	.140	.064	.159	.087	.023
		(iii)	.171	.103	.029	.141	.080	.016

NOTE: See the Note for Table 3.

[a] ρ is the correlation coefficient between the gap times U and V. U and V are generated from the distributions of Gumbel (1960) if $\rho = 0$ or .25, and from the distributions of Lawrance and Lewis (1981) if $\rho = .50$.

[b] About 32% of the first failure times are censored and about 57% of the second failure times are censored.

if the ith patient was in the high-dose group, and 0 otherwise. The corresponding regression coefficients β_{1k} and β_{2k} can be interpreted, respectively, as the treatment effects from the low-dose and high-dose ribavirin after k months of treatment.

For the data in Table 1, the parameter estimates $\hat{\beta}_{11}$, $\hat{\beta}_{21}$, $\hat{\beta}_{12}$, $\hat{\beta}_{22}$, $\hat{\beta}_{13}$, and $\hat{\beta}_{23}$ are, respectively, -1.394, $-.938$, $-.655$, $.020$, $-.615$, and $-.331$. The corresponding covariance matrix estimate is

$$\hat{Q} = \begin{bmatrix} .245 & .075 & .051 & .017 & .107 & .061 \\ & .136 & .026 & .041 & .046 & .091 \\ & & .287 & .119 & .133 & .091 \\ & & & .167 & .076 & .069 \\ & & & & .257 & .114 \\ & & & & & .229 \end{bmatrix}. \quad (3.1)$$

For testing $H_{lk} : \beta_{lk} = 0$ ($l = 1, 2$; $k = 1, 2, 3$) jointly, the observed value of the quadratic form W is 13.058 with 6 df, which is significant at the 5% level. This result, however, does not indicate which β_{lk}'s are nonzero. On the other hand, the sequential multiple testing procedure described in Section 2 can provide an overall picture about the treatment effects over time. Now, suppose that the drug ribavirin is no worse than the placebo, that is, $\beta_{lk} \leq 0$ ($l = 1, 2$; $k = 1, 2, 3$), and we would like to know which β_{lk}'s are strictly negative. Here, the null hypotheses are H_{lk}, and the alternative hypotheses are $H^a_{lk} : \beta_{lk} < 0$ ($l = 1, 2$; $k = 1, 2, 3$). The standardized parameter estimates $\tilde{\beta}_{11}$, $\tilde{\beta}_{21}$, $\tilde{\beta}_{12}$, $\tilde{\beta}_{22}$, $\tilde{\beta}_{13}$, and $\tilde{\beta}_{23}$ are, respectively, -2.814, -2.546, -1.224, $.049$, -1.213, and $-.691$. The corresponding covariance matrix estimate $\tilde{Q}$ can be obtained from $\hat{Q}$ in (3.1). Now, let $(V_1, \ldots, V_6)'$ be a multivariate normal random vector with mean 0 and covariance matrix $\tilde{Q}$. Since $\tilde{\beta}_{11} < \tilde{\beta}_{21} < \tilde{\beta}_{12} < \tilde{\beta}_{13} < \tilde{\beta}_{23} < \tilde{\beta}_{22}$, we will test the null hypotheses in the sequence of H_{11}, H_{21}, H_{12}, H_{13}, H_{23}, and H_{22}. Using the integration algorithm for multivariate normal probabilities provided by Schervish (1984), we have

$$\Pr(\min\{V_1, V_2, V_3, V_4, V_5, V_6\} \leq -2.814) \approx .013,$$

$$\Pr(\min\{V_2, V_3, V_4, V_5, V_6\} \leq -2.546) \approx .024,$$

and

$$\Pr(\min\{V_3, V_4, V_5, V_6\} \leq -1.224) \approx .294.$$

These results indicate that the drug ribavirin (either in low dose or high dose) prolonged the time to virus positivity in the beginning of the trial. There is no strong evidence, however, that this favorable drug effect extended throughout the entire study. The combined estimates of drug effects across three months for the low-dose and high-dose groups are $-.972$ and $-.502$, respectively, with the corresponding estimated standard errors .386 and .306. It is interesting to observe that the results from the low-dose group appear to be more significant than those from the high-dose group.

For the second example, we consider the tumor recurrence data for bladder cancer patients. Here, $n = 86$ and $K = 4$. The failure time $\tilde{X}_{ki}$ is the number of months from the beginning of the treatment to the kth tumor recurrence for the ith patient. Let $Z_{ki} = (Z_{1ki}, Z_{2ki}, Z_{3ki})'$, where $Z_{1ki} = 1$ if the ith patient was in the thiotepa group and 0 if the ith patient was in the control group, and Z_{2ki} and Z_{3ki} are, respectively, the number of initial tumors and the size of the largest initial tumor for the ith patient. Here, we are mainly interested in the treatment effects, $\eta_k = \beta_{1k}$ ($k = 1, 2, 3, 4$).

The estimates $\hat{\eta}_k$'s are shown in the first row of Table 5. The variance–covariance estimate of $(\hat{\eta}_1, \ldots, \hat{\eta}_4)'$ is

$$\hat{\Psi} = \begin{bmatrix} .095 & .060 & .057 & .044 \\ & .132 & .130 & .116 \\ & & .172 & .159 \\ & & & .240 \end{bmatrix}. \quad (3.2)$$

The observed value of the quadratic form W for testing jointly $H_k : \eta_k = 0$ $(k = 1, 2, 3, 4)$ is 3.967 with 4 df, which does not indicate that any of the η_k's are nonzero.

Now, let us assume that $\eta_k \leq 0$. The standardized estimates of treatment effects $\tilde{\eta}_1$, $\tilde{\eta}_2$, $\tilde{\eta}_3$, and $\tilde{\eta}_4$ are -1.683, -1.702, -1.686, and -1.329, respectively. Let $(V_1, \ldots, V_4)'$ be a zero-mean normal random vector with covariance matrix $\tilde{\Psi}$, which is the correlation matrix based on (3.2). Then

$$\Pr(\min\{V_1, V_2, V_3, V_4\} \leq -1.702) \approx .115,$$

$$\Pr(\min\{V_1, V_3, V_4\} \leq -1.686) \approx .105,$$

$$\Pr(\min\{V_1, V_4\} \leq -1.683) \approx .086,$$

and

$$\Pr(V_4 \leq -1.329) \approx .092.$$

These results provide some evidence, though not very strong, that thiotepa slows down tumor recurrences. On the other hand, the combined estimate $\hat{\eta}$ is $-.549$ with estimated standard error .285, which provides stronger evidence for a beneficial treatment effect.

For comparisons, we also analyzed the tumor recurrence data by the methods proposed by Prentice et al. (1981). For their procedures, the patients who were censored before the kth tumor recurrences are excluded from the risk sets for the $(k + 1)$th recurrences. Thus their risk sets are smaller than ours (see Fig. 1). The second and third rows in the body of Table 5 display the results from models (2) and (3) in Prentice et al. (1981). Notice that for the first recurrence the standard error estimate based on model (2) or model (3) is different from that of the new method because the former is based on the matrix of the second derivatives of the log partial likelihood, whereas the latter is obtained from $\hat{Q}$ in (A.3), which is in fact a robust estimate (see Lin and Wei 1989). It is difficult to interpret some of the results from the Prentice et al. methods in Table 5. For example, based on model (2), treatment effects appear to be significant with respect to the first and third recurrences but not to the second recurrence. This difficulty probably arises from the fact that the sizes of the risk sets defined in their procedures are rather small beyond the first recurrence (see Fig. 1). For completeness, the tumor recurrence data are analyzed by the method of Andersen and Gill (1982). The maximum partial likelihood estimate of treatment effect with adjustments for the other two covariates is $-.407$, with estimated standard error .200.

4. REMARKS

The proposed semiparametric methods and related inferential procedures provide some conceptually straightforward approaches to the analysis of general multivariate failure time data. They avoid modeling the structure of dependence among distinct types of failure times within each subject. Moreover, they are applicable to moderate-sized samples. Finally, the parameter estimates $\hat{\beta}_k$'s can be obtained from an existing statistical package, and the covariance matrix estimate $\hat{Q}$ can be easily computed through a simple FORTRAN program, which is available from the authors.

Recently, Hougaard (1986) proposed a rich class of continuous multivariate lifetime distributions. The dependence between individuals in a group is modeled by a group-specific quantity, which is treated as an unobserved covariate Z common to the individuals in the group and assumed to follow a positive stable distribution. With this approach, the marginal distributions do not identify the dependence parameter. If one assumes, however, that Z follows a gamma distribution (see Clayton and Cuzick 1985), then the dependence parameter and the regression parameter are confounded. The implementation of the latter approach tends to be more complicated than the one studied by Hougaard (1986). Although Hougaard's model seems rather promising for practical use, the statistical properties of some of the inference procedures are still unknown. Furthermore, for the recurrent events, Hougaard's model

Table 5. Proportional Hazards Analyses of Treatment Effects on Tumor Recurrences of Bladder Cancer Patients With Adjustments for the Number and Size of Initial Tumors

	Recurrence number				All recurrences
Model	1: $\hat{\eta}_1$	2: $\hat{\eta}_2$	3: $\hat{\eta}_3$	4: $\hat{\eta}_4$	$(\eta_1 = \cdots = \eta_4 = \eta)$ $\hat{\eta}$
New method	.518	.619	.700	.651	−.549
	(.308)	(.364)	(.415)	(.490)	(.285)
PWP[a] model (2)	−.518	−.426	−.899	−.237	−.490
	(.316)	(.402)	(.540)	(.683)	(.209)
PWP model (3)	−.518	−.259	.221	−.195	−.270
	(.316)	(.405)	(.549)	(.642)	(.208)
AG[b] model	—	—	—	—	−.407
	—	—	—	—	(.200)

NOTE: Estimated standard errors are in parentheses.
[a] PWP denotes Prentice, Williams, and Peterson (1981).
[b] AG denotes Andersen and Gill (1982).

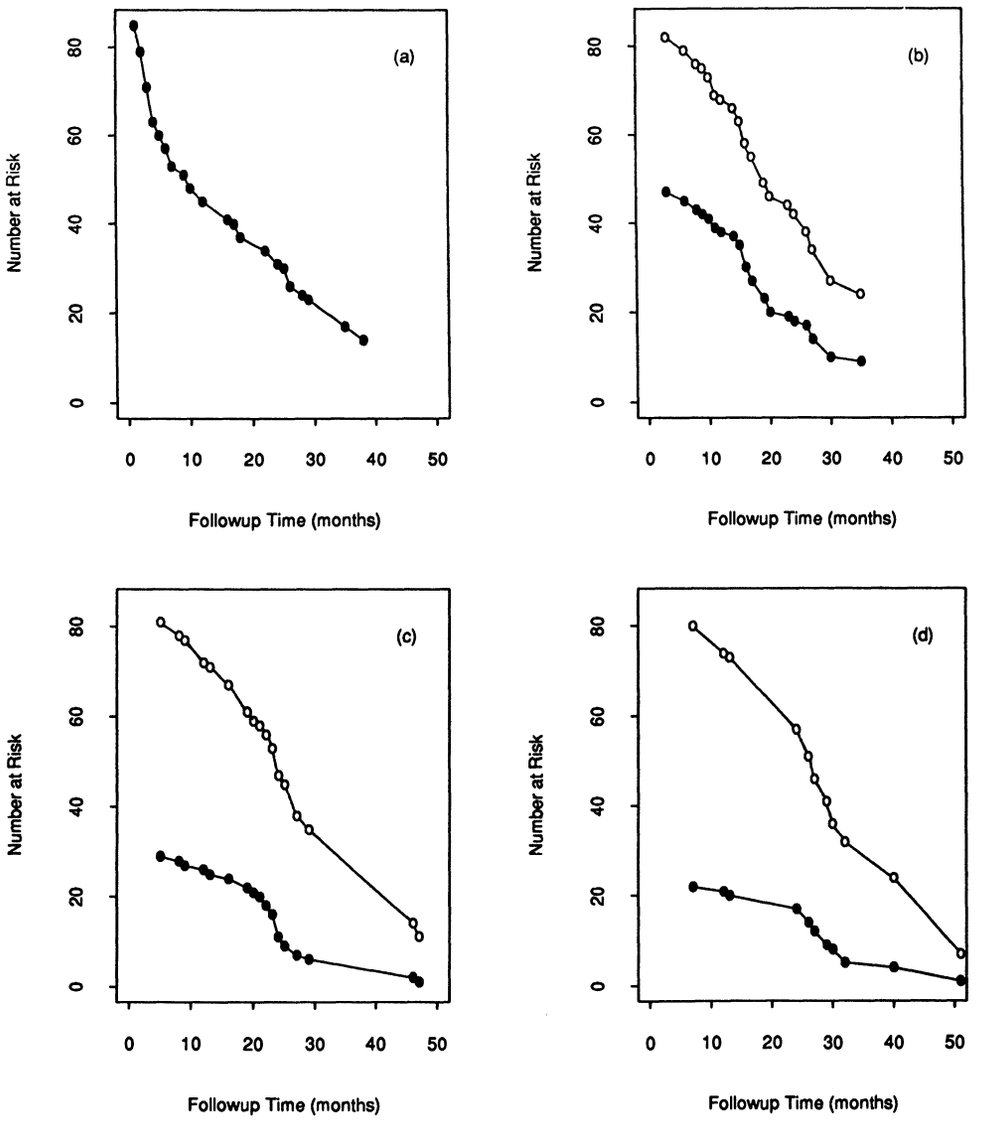

Figure 1. The Sizes of the Risk Sets in the Analyses of the Tumor Recurrence Data by the New Method (○) and the Prentice et al. Model (2) (●): (a) First Recurrence (identical for the two methods); (b) Second Recurrence; (c) Third Recurrence; (d) Fourth Recurrence.

has difficulty incorporating the natural ordering among distinct recurrence times. It is interesting to note that when using his procedure for the proportional hazards model, Hougaard (1986) suggested that one estimate the marginal hazards and their regression parameters based on the univariate theory and then compute the maximum likelihood estimate of the dependence parameter by restricting the marginals to the estimated ones. If the aim is only to estimate the effect of a treatment, say, in the presence of dependence, there seems to be no gain in efficiency by using this so-called two-stage procedure.

APPENDIX: ASYMPTOTIC PROPERTIES OF PARAMETER ESTIMATORS

We show here that for large n the distribution of the random vector $(\hat{\beta}_1', \ldots, \hat{\beta}_K')'$ can be approximated by a pK-dimensional normal distribution with mean vector $(\beta_1', \ldots, \beta_K')'$ and with a covariance matrix that can be easily estimated. For the kth type of failure, let

$$N_{ki}(t) = I\{X_{ki} \le t, \Delta_{ki} = 1\},$$

$$Y_{ki}(t) = I\{X_{ki} \ge t\},$$

and

$$M_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(u)\lambda_{ki}(u)\, du,$$

where $I\{\cdot\}$ is the indicator function. In addition, let $C_k(\beta; t)$ be the logarithm of the Cox likelihood (2.2) for the kth type of failure evaluated at time t. Then, we have

$$C_k(\beta; t) = \sum_{i=1}^n \int_0^t \beta' Z_{ki}(u) dN_{ki}(u) - \int_0^t \log\left[\sum_{i=1}^n Y_{ki}(u)\exp\{\beta' Z_{ki}(u)\}\right] d\bar{N}_k(u),$$

where $\bar{N}_k(u) = \sum_{i=1}^n N_{ki}(u)$. The vector of derivatives $U_k(\beta; t)$ of $C_k(\beta; t)$ with respect to β has the form

$$U_k(\beta; t) = \sum_{i=1}^n \int_0^t Z_{ki}(u) dN_{ki}(u) - \int_0^t \frac{\sum_{i=1}^n Y_{ki}(u) Z_{ki}(u)\exp\{\beta' Z_{ki}(u)\}}{\sum_{i=1}^n Y_{ki}(u)\exp\{\beta' Z_{ki}(u)\}} d\bar{N}_k(u).$$

It follows immediately that

$$U_k(\beta_k; t) = \sum_{i=1}^{n} \int_0^t Z_{ki}(u)dM_{ki}(u) - \int_0^t \frac{\sum_{i=1}^n Y_{ki}(u)Z_{ki}(u)\exp\{\beta_k' Z_{ki}(u)\}}{\sum_{i=1}^n Y_{ki}(u)\exp\{\beta_k' Z_{ki}(u)\}} d\bar{M}_k(u),$$

where $\bar{M}_k(u) = \sum_{i=1}^n M_{ki}(u)$.

By the Taylor series expansion of $U_k(\beta_k; \infty)$ around β_k, we have

$$n^{-1/2}U_k(\beta_k; \infty) = \hat{A}_k(\beta_k^*)n^{1/2}(\hat{\beta}_k - \beta_k), \quad \text{(A.1)}$$

where

$$\hat{A}_k(\beta) = n^{-1}\sum_{j=1}^{n} \Delta_{kj}\left[\frac{\sum_{i=1}^n Y_{ki}(X_{kj})Z_{ki}(X_{kj})^{\otimes 2}\exp\{\beta' Z_{ki}(X_{kj})\}}{\sum_{i=1}^n Y_{ki}(X_{kj})\exp\{\beta' Z_{ki}(X_{kj})\}} - \left(\frac{\sum_{i=1}^n Y_{ki}(X_{kj})Z_{ki}(X_{kj})\exp\{\beta' Z_{ki}(X_{kj})\}}{\sum_{i=1}^n Y_{ki}(X_{kj})\exp\{\beta' Z_{ki}(X_{kj})\}}\right)^{\otimes 2}\right].$$

Here, $a^{\otimes 2}$ denotes the matrix aa' for a column vector a, and β_k^* is on the line segment between $\hat{\beta}_k$ and β_k. By theorem 4.2 of Andersen and Gill (1982), the matrix $\hat{A}_k(\beta_k^*)$ converges in probability to a nonsingular deterministic matrix, denoted by $A_k(\beta_k)$, which can be consistently estimated by $\hat{A}_k(\hat{\beta}_k)$. It is important to note that although $U_k(\beta_k; t)$ is a local square integrable martingale in t with respect to the kth type of failure, the asymptotic joint distribution of $n^{-1/2}U_1(\beta_1; \infty)$, . . . and $n^{-1/2}U_K(\beta_K; \infty)$ cannot be derived directly from Rebolledo's central limit theorem for local square integrable martingales.

Now, let

$$S_k^{(1)}(\beta; t) = n^{-1}\sum_{i=1}^{n} Y_{ki}(t)Z_{ki}(t)\exp\{\beta' Z_{ki}(t)\},$$

$$S_k^{(0)}(\beta; t) = n^{-1}\sum_{i=1}^{n} Y_{ki}(t)\exp\{\beta' Z_{ki}(t)\},$$

$$s_k^{(1)}(\beta; t) = E[Y_{k1}(t)Z_{k1}(t)\exp\{\beta' Z_{k1}(t)\}],$$

and

$$s_k^{(0)}(\beta; t) = E[Y_{k1}(t)\exp\{\beta' Z_{k1}(t)\}].$$

Then, using arguments similar to those given in the proofs of theorem 4.2 of Andersen and Gill (1982) and of theorems 4.2.1 and 4.3.1 of Gill (1980), one can show that for any β,

$$n^{-1/2}\int_0^\infty \left\{\frac{S_k^{(1)}(\beta; t)}{S_k^{(0)}(\beta; t)} - \frac{s_k^{(1)}(\beta; t)}{s_k^{(0)}(\beta; t)}\right\} d\bar{M}_k(t) \to 0$$

in probability as $n \to \infty$. Therefore, $n^{-1/2}U_k(\beta_k; \infty)$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\left\{\int_0^\infty Z_{ki}(t)dM_{ki}(t) - \int_0^\infty \frac{s_k^{(1)}(\beta_k; t)}{s_k^{(0)}(\beta_k; t)} dM_{ki}(t)\right\},$$

which is a sum of n iid random vectors. It follows from the multivariate central limit theorem and (A.1) that $n^{1/2}(\hat{\beta}_1' - \beta_1', \ldots, \hat{\beta}_K' - \beta_K')'$ converges in distribution to a zero-mean normal random vector. In addition, the asymptotic covariance matrix between $n^{1/2}(\hat{\beta}_k - \beta_k)$ and $n^{1/2}(\hat{\beta}_l - \beta_l)$ is given by

$$D_{kl}(\beta_k, \beta_l) = A_k^{-1}(\beta_k)E\{w_{k1}(\beta_k)w_{l1}(\beta_l)'\}A_l^{-1}(\beta_l),$$

where $w_{kj}(\beta_k) = \int_0^\infty \{Z_{kj}(t) - s_k^{(1)}(\beta_k; t)/s_k^{(0)}(\beta_k; t)\} dM_{kj}(t)$. It is natural to estimate $E\{w_{k1}(\beta_k)w_{l1}(\beta_l)'\}$ by

$$\hat{B}_{kl}(\hat{\beta}_k, \hat{\beta}_l) = n^{-1}\sum_{j=1}^{n} W_{kj}(\hat{\beta}_k)W_{lj}(\hat{\beta}_l)', \quad \text{(A.2)}$$

where

$$W_{kj}(\beta_k) = \Delta_{kj}\left\{Z_{kj}(X_{kj}) - \frac{S_k^{(1)}(\beta_k; X_{kj})}{S_k^{(0)}(\beta_k; X_{kj})}\right\} - \sum_{m=1}^{n} \frac{\Delta_{km}Y_{kj}(X_{km})\exp\{\beta_k' Z_{kj}(X_{km})\}}{nS_k^{(0)}(\beta_k; X_{km})} \times \left\{Z_{kj}(X_{km}) - \frac{S_k^{(1)}(\beta_k; X_{km})}{S_k^{(0)}(\beta_k; X_{km})}\right\}.$$

Notice that $W_{kj}(\beta_k)$ is obtained by substituting $s_k^{(1)}(\beta_k; t)$, $s_k^{(0)}(\beta_k; t)$, and $\lambda_{k0}(t)dt$ with $S_k^{(1)}(\beta_k; t)$, $S_k^{(0)}(\beta_k; t)$, and $nd\bar{N}_k(t)/S_k^{(0)}(\beta_k; t)$, respectively, in $w_{kj}(\beta_k)$. The proof of the consistency of the estimator (A.2) is rather tedious. However, it essentially uses the same techniques as in the proofs of theorem 1 of Wei and Lachin (1984) and of theorem 3.2 and corollary 3.3 of Andersen and Gill (1982). It follows that the asymptotic covariance matrix between $n^{1/2}(\hat{\beta}_k - \beta_k)$ and $n^{1/2}(\hat{\beta}_l - \beta_l)$ can be consistently estimated by

$$\hat{D}_{kl}(\hat{\beta}_k, \hat{\beta}_l) = \hat{A}_k^{-1}(\hat{\beta}_k)\hat{B}_{kl}(\hat{\beta}_k, \hat{\beta}_l)\hat{A}_l^{-1}(\hat{\beta}_l).$$

Finally, for large n, the covariance matrix of $(\hat{\beta}_1', \ldots, \hat{\beta}_K')'$ can be estimated by

$$\hat{Q} = n^{-1}\begin{bmatrix} \hat{D}_{11}(\hat{\beta}_1, \hat{\beta}_1) & \cdots & \hat{D}_{1K}(\hat{\beta}_1, \hat{\beta}_K) \\ \vdots & & \vdots \\ \hat{D}_{K1}(\hat{\beta}_K, \hat{\beta}_1) & \cdots & \hat{D}_{KK}(\hat{\beta}_K, \hat{\beta}_K) \end{bmatrix}. \quad \text{(A.3)}$$

[*Received November 1987. Revised April 1989.*]

REFERENCES

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100–1120.

Andrews, D. F., and Herzberg, A. M. (1985), *Data: A Collection of Problems From Many Fields for the Student and Research Worker*, New York: Springer-Verlag.

Byar, D. P. (1980), "The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine, and Topical Thiotepa," in *Bladder Tumors and Other Topics in Urological Oncology*, eds. M. Pavone-Macaluso, P. H. Smith, and F. Edsmyn, New York: Plenum, pp. 363–370.

Clayton, D., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 148, 82–117.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

——— (1975), "Partial Likelihood," *Biometrika*, 62, 269–276.

Cox, D. R., and Oakes, D. A. (1984), *Analysis of Survival Data*, New York: Chapman & Hall.

Gail, M. H., Santner, T. J., and Brown, C. C. (1980), "An Analysis of Comparative Carcinogenesis Experiments Based on Multiple Times to Tumor," *Biometrics*, 36, 255–266.

Gill, R. D. (1980), *Censoring and Stochastic Integrals* (Tract 124), Amsterdam: Mathematical Centre.

Gumbel, E. J. (1960), "Bivariate Exponential Distributions," *Journal of the American Statistical Association*, 55, 698–707.

Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6, 65–70.

Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671–678.

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

Lawless, J. F. (1987), "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association*, 82, 808–815.

Lawrance, A. J., and Lewis, P. A. W. (1981), "A New Autoregressive Time Series Model in Exponential Variables," *Advances in Applied Probability*, 13, 826–845.

Lin, D. Y., and Wei, L. J. (1989), "The Robust Inference for the Cox Proportional Hazards Model," *Journal of the American Statistical Association*, 84, 1074–1078.

Makuch, R. W., and Parks, W. P. (1988), "Statistical Methods for the Analysis of HIV-1 Core Polypeptide Antigen Data in Clinical Studies," *AIDS Research and Human Retroviruses*, 4, 305–316.

Marcus, R., Peritz, E., and Gabriel, K. R. (1976), "On Closed Testing Procedures With Special Reference to Ordered Analysis of Variance," *Biometrika*, 63, 655–660.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981), "On the Regression Analysis of Multivariate Failure Time Data," *Biometrika*, 68, 373–379.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge, U.K.: Cambridge University Press.

Schervish, M. J. (1984), "Multivariate Normal Probabilities With Error Bound," *Applied Statistics*, 33, 81–94; Correction (1985), 34, 103–104.

Wei, L. J., and Johnson, W. E. (1985), "Combining Dependent Tests With Incomplete Repeated Measurements," *Biometrika*, 72, 359–364.

Wei, L. J., and Lachin, J. M. (1984), "Two-Sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations," *Journal of the American Statistical Association*, 79, 653–661.

Wei, L. J., and Stram, D. O. (1988), "Analyzing Repeated Measurements With Possibly Missing Observations by Modelling Marginal Distributions," *Statistics in Medicine*, 7, 139–148.