STATISTICS IN MEDICINE
Statist. Med. **17**, 999–1015 (1998)

PARAMETRIC LIKELIHOODS FOR MULTIPLE NON-FATAL COMPETING RISKS AND DEATH

YU SHEN* AND PETER F. THALL

Department of Biomathematics, Box 237, UT M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.

SUMMARY

Clinical trials of fatal diseases often focus on one or more non-fatal events, in addition to survival, both to characterize morbidity and to improve survival estimates. Three statistical complications are that the time to each non-fatal event and subsequent residual survival may be either positively or negatively associated, the times to death with or without an antecedent event often have very different distributions, and death may censor some of the non-fatal event times. Consequently, the overall survival time distribution is a mixture of the distributions corresponding to the possible antecedent non-fatal events. These conditions violate the usual assumptions underlying many statistical methods for analysing multivariate time-to-event data. In this paper, we consider a general parametric model for multiple non-fatal competing risks and death. The model accounts for positive or negative association between the time of each non-fatal event and subsequent survival while accommodating covariates and the usual administrative censoring. Each event time distribution is specified marginally by a three-parameter generalized odds rate model, and the time of each non-fatal event and subsequent residual survival are combined under a bivariate generalized von Morgenstern distribution. The approach is illustrated by application to two data sets from clinical trials in colon cancer and acute leukaemia. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

In clinical trials of treatments for fatal diseases, the times to one or more non-fatal events may be recorded to characterize morbidity *per se* and also provide auxiliary information that may improve estimates of overall and residual survival.[1–6] A prototypical example of one non-fatal event and death is a cancer trial in which remission duration (disease progression, time to relapse) and survival time are both clinically and scientifically important. Aside from administrative censoring, patients may either relapse and die, or die without antecedent relapse, that is, suffer 'regimen-related' or 'progression-free' death. An even more complex data structure arises in trials of chemotherapeutic regimens for acute leukaemia in which each patient is either brought into complete remission (CR) or declared resistant and removed from the study so that a different therapeutic approach may be taken. The times to being declared in CR or resistant are non-fatal

* Correspondence to: Y. Shen, Department of Biomathematics, Box 237, UT M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.

CCC 0277–6715/98/09999–17$17.50

Received September 1996
Revised April 1997

competing risks, and each is closely associated with subsequent residual survival time. Patients who achieve CR quickly are likely to survive longer, hence the problem of characterizing negative association arises. Finally, a patient may die during the complex process of attempted remission induction, often due to adverse effects of the treatment regimen itself, before either non-fatal event has occurred.

The scientific goals in such medical settings typically include evaluation of treatment effects and estimation of morbidity and mortality, both overall and for individual patients or patient subgroups. These goals in turn require estimation of the distributions of the non-fatal event times, of residual survival given a particular non-fatal event at a specified time, and of overall survival, each distribution possibly adjusted for covariate effects. For example, it is very important prognostically to estimate the residual survival time of a cancer patient given his or her relapse time.

In this paper, we consider the general setting in which each patient may experience at most one of several non-fatal events, that is, in which the events are competing risks, while death may occur before any of the non-fatal events. The three complications motivating our approach are: (i) death censors all non-fatal events that have not occurred; (ii) the time to a non-fatal event and subsequent residual survival time may be either positively or negatively associated, and (iii) the overall survival distribution is a mixture of two or more different distributions corresponding to death either with or without an antecedent non-fatal event. With these considerations and the array of statistical inferences previously described in mind, we employ a family of parametric models sufficiently general to allow broad application but that lends itself readily to interpretation and explanation to medical colleagues. We use a three-parameter generalized odds rate model[7] for the marginal distributions of the time to each non-fatal event, the residual survival time after a non-fatal event, and the time to death without any antecedent event. We assume that the times to the non-fatal events and event-free death are mutually independent.[8] For the bivariate distribution of the time to each non-fatal event and subsequent residual survival, we use a bivariate generalized von Morgenstern distribution,[9] which characterizes the positive or negative association between these two times by a single parameter. The model accommodates one or more non-fatal competing risks and death, covariates with effects specific to each marginal time, and the usual administrative censoring. In the case of one non-fatal event, such as relapse, the model is a generalization of that proposed by Lagakos,[1,2] who addressed the problem of using the observed time of a non-fatal event as an *auxiliary variable* to improve estimation of survival.

The remainder to the paper is organized as follows. In Section 2 we define formally the general data structure and model. In Section 3 we consider the case of one non-fatal event and death, including analysis of the colon cancer data previously analysed by Moertel *et al.*[10] and Lin.[11] In Section 4, we apply the general model to a data set arising from chemotherapy of acute leukaemia, in which the times to CR or resistance are non-fatal competing risks.

2. DATA STRUCTURE AND MODEL

Let T_0 be the time to death without any antecedent non-fatal event, $T_{j,1}$ the time to the jth non-fatal event, and $T_{j,2}$ the subsequent residual survival time, $j = 1, \ldots, K$. The non-fatal events and event-free death are *competing risks* in that at most one of $T_0, T_{1,1}, \ldots, T_{K,1}$ may be observed. We denote $\mathbf{T}_j = (T_{j,1}, T_{j,2})$ and $T = \min\{T_{1,1}, \ldots, T_{K,1}\}$. Aside from right (administrative) censoring, one observes either $\mathbf{T}_j$ if $T_{j,1} = T < T_0$, or T_0 alone if $T > T_0$.

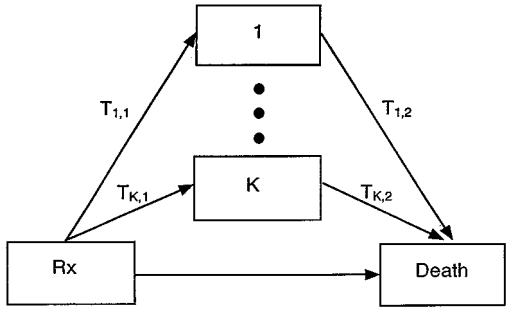

Figure 1. K non-fatal competing risks and death

Table I. Possible patient outcomes, death and K non-fatal competing risks

Observed data	δ_0	$\delta_1, \ldots, \delta_K$
$T_{j,1}, T_{j,2}, I[T = T_{j,1} < T_0, T_{j,1} + T_{j,2} < U]$	1	$\delta_j = 1, \delta_k = 0, k \neq j$
$T_{j,1}, U, I[T = T_{j,1} < T_0, T_{j,1} < U < T_{j,1} + T_{j,2}]$	0	$\delta_j = 1, \delta_k = 0, k \neq j$
$T_0, I[T_0 < \min\{T, U\}]$	1	$\delta_1 = \cdots = \delta_K = 0$
$U, I[U < \min\{T, T_0\}]$	0	$\delta_1 = \cdots = \delta_K = 0$

In the first case the patient's survival time is $T_{j,1} + T_{j,2}$, while in the second case it is T_0. As noted earlier, one motivation for the model proposed here is that the survival times T_0 and $T_{j,1} + T_{j,2}, 1 = 1, \ldots, K$ may have very different distributions. In particular, overall survival time is the mixture

$$T_D = T_0 I[T_0 < T] + \sum_{j=1}^{K} (T_{j,1} + T_{j,2}) I[T < T_0] I[T = T_{j,1}] \tag{1}$$

where $I[A]$ denotes the indicator of the event A. Figure 1 illustrates the set of possible outcomes for each patient.

To account additionally for non-informative right censoring at time U, we define the following indicators. For the jth non-fatal event, let $\delta_j = 1$ if $T_{j,1} = T < \min\{T_0, U\}$ and 0 otherwise. Let $\delta_0 = 1$ if the time of death is observed, formally if either $T_0 < \min\{T, U\}$ or $T_0 > T = T_{j,1}$ and $T_{j,1} + T_{j,2} < U$, with $\delta_0 = 0$ otherwise. We can regard all of these random variables as latent event times, since each may or may not be observed depending on which $T_{j,1} = T$, whether $T < T_0$, and the value of U. Table I gives the four forms that a patient's observed outcomes may take.

We assume that $T_0, T_{1,1}, \ldots, T_{K,1}$ are mutually independent, because the available data do not allow us to distinguish between this and a model that accounts for dependence among these competing risks. That is, it is not possible to verify empirically whether the assumption of independence among competing risks is true. Discussions of this problem and associated identifiability problems in the context of competing risks are given by Gail,[12] Prentice *et al.*[13] and Kalbfleisch and Prentice.[8]

Given this assumption and the above data structure, our modelling strategy is first to specify parametric distributions for the marginals of T_0 and each $T_{j,r}$, $r = 1, 2$ and then a bivariate distribution for each pair $(T_{j,1}, T_{j,2})$ having their given marginals. Our criteria for the bivariate distribution are that it must be a tractable, parsimonious model that allows either positive or negative association. First, denote the marginal CDF, survivor function, and probability density function of the kth event time by F_k, $S_k = 1 - F_k$, and $f_k = F_k'$, respectively, where the index k denotes 0 or any of the double indices $(j, 1)$ or $(j, 2)$. We use the following parametric bivariate distribution for each $\mathbf{T}_j$, apparently first introduced by von Morgenstern.[9] For each $j = 1, \ldots, K$,

$$F_{j,12}(t_1, t_2) = \Pr[T_{j,1} \leqslant t_1, T_{j,2} \leqslant t_2] = F_{j,1}(t_1)F_{j,2}(t_2)[1 + \alpha_j S_{j,1}(t)S_{j,2}(t_2)], \quad t_1, t_2 > 0 \quad (2)$$

$-1 < \alpha_j < +1$. This distribution is characterized by its marginals $F_{j,1}$ and $F_{j,2}$ and the dependence parameter α_j, with $\alpha_j = 0$ corresponding to independence between $T_{j,1}$ and $T_{j,2}$. This model was studied in the case of Weibull marginals by Butkiewicz and Hys,[14] and also discussed by Johnson and Kotz.[15] The generalized odds rate and von Morgenstern distributions are not essential to our approach, however. In general, one may use any tractable marginal and bivariate distributions with the above properties that are appropriate for a given application.

We may now write down the general form of a single patient's likelihood, which is the product of the following four terms corresponding to the rows of Table I. Recall that at most one of the indicators $\delta_1, \ldots, \delta_K$ may be non-zero and that δ_0 is the indicator that the time of death is observed, regardless of whether death is preceded by a non-fatal event:

$$\mathscr{L}_1 = \prod_{j=1}^{K} \left[f_{j,12}(T_1, T_2)S_0(T_1) \prod_{k \neq j} S_{k,1}(T_1) \right]^{\delta_0 \delta_j} \quad (3)$$

$$\mathscr{L}_2 = \prod_{j=1}^{K} \left[\int_{t_2 = U - T_1}^{\infty} f_{j,12}(T_1, t_2)\, \mathrm{d}t_2 S_0(T_1) \prod_{k \neq j} S_{k,1}(T_1) \right]^{(1-\delta_0)\delta_j} \quad (4)$$

$$\mathscr{L}_3 = \left[f_0(T_0) \prod_{j=1}^{K} S_{j,1}(T_0)^{(1-\delta_j)} \right]^{\delta_0} \quad (5)$$

$$\mathscr{L}_4 = \left[S_0(U) \prod_{j=1}^{K} S_{j,1}(U)^{(1-\delta_j)} \right]^{(1-\delta_0)}. \quad (6)$$

The full likelihood is obtained as usual by forming the product over patients of the individual likelihoods $\mathscr{L} = \mathscr{L}_1\mathscr{L}_2\mathscr{L}_3\mathscr{L}_4$.

In theory any marginal distributions for non-negative-valued random variables appropriate for the particular application may be used for F_0 and each $F_{j,r}$. We employ the three-parameter generalized odds rate model, characterized by the survival function $S(t; \lambda, \phi, c) = [1 + c(t/\lambda)^{\phi}]^{-1/c}$, $\lambda > 0$, $\phi > 0$, $c > 0$. This is an especially flexible and tractable family of distributions.[7,17] It contains the log-logistic distribution $S(t; \lambda, \phi, 1) = [1 + (t/\lambda)^{\phi}]^{-1}$ when $c = 1$, discussed by Bennett[16] and has the Weibull as the limiting case obtained as $S(t; \lambda, \phi, 0) =_{\mathrm{def}} \lim_{c \downarrow 0} S(t; \lambda, \phi, c) = \exp(-t/\lambda)^{\phi}$. Patient covariates $\mathbf{Z} = (Z_1, \ldots, Z_p)$ may be incorporated into the model in terms of the linear component $\beta'\mathbf{Z} = \beta_1 Z_1 + \cdots + \beta_p Z_p$ by defining the more general form

$$S(t; \mathbf{Z}, c, \phi, \lambda, \boldsymbol{\beta}) = [1 + c(t/\lambda)^{\phi} \mathrm{e}^{\beta'\mathbf{z}}]^{-1/c}. \quad (7)$$

To interpret covariate effects under this model, consider the two-sample case where Z is the treatment group indicator and $\boldsymbol{\beta}'\mathbf{Z} = \beta Z$. In the case of proportional hazards ($c = 0$), β is the log relative risk between the two groups. Under the proportional odds case ($c = 1$), β is the log of the odds ratio of survival beyond a given time. For arbitrary $c > 0$, β is the *generalized log odds ratio* $\log[\{(1 - S_1(t)^c)/S_1(t)^c\}/\{(1 - S_0(t)^c)/S_0(t)^c\}]$, where $S_0(t) = S(t \mid Z = 0)$, and $S_1(t) = S(t \mid Z = 1)$. Discussions are given by Dabrowska and Doksum,[7] Clayton and Cuzick[17] and Wingo.[18] While we may use the alternative parameterization $[1 + c(t/e^{\beta'\mathbf{Z}})^{\phi}]^{-1/c}$, when $\phi \neq 1$ the above interpretation of covariate effects does not apply.

We denote $\boldsymbol{\theta}_k = (\lambda_k, \phi_k, c_k, \beta_{k,1}, \ldots, \beta_{k,p})$, so that the model is parameterized by $\alpha_1, \ldots, \alpha_k$ and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_k, k = 0, (1,1), (1,2), \ldots, (K,1), (K,2)\}$. For each application, we obtained maximum likelihood estimates (MLEs) under the constraints that all $\lambda_k, \phi_k, c_k > 0$ and $-1 < \alpha_k < +1$ for all j, k using the method of Gay,[19] which in particular does not require specification of any derivatives of the likelihood. We estimate standard errors of the MLEs and confidence bands for median residual survival using bootstrapping. While the traditional approach to maximum likelihood is to use Fisher scoring to obtain MLEs and the information matrix to estimate standard errors, by relying on the optimization method of Gay and bootstrapping (Efron and Tibshirani[20]) we avoid computation of derivatives of the log-likelihood, which can be rather complicated when $K \geqslant 2$. Our simulation results, not presented here, indicate that bootstrap samples of size 200 provide standard error estimates very close to the sample standard errors obtained from 1000 simulations.

3. ONE NON-FATAL EVENT AND DEATH

3.1. Model

We first consider the case of a single non-fatal event such as disease progression or relapse. Since $K = 1$ we drop the indices j, k. From (1), the bivariate CDF for (T_1, T_2) is

$$F_{12}(t_1, t_2) = \Pr[T_1 \leqslant t_1, T_2 \leqslant t_2] = F_1(t_1)F_2(t_2)\{1 + \alpha S_1(t_1)S_2(t_2)\},\ 0 < t_1, t_2 < \infty. \quad (8)$$

Aside from $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, the model has ten parameters with (ϕ_j, λ_j, c_j) characterizing the marginal $F_j, j = 0, 1, 2$ and α the association between T_1 and T_2. In this case, the possible forms that each patient's outcome data may take, specified generally in Table I, are given in Table II.

The likelihood function given in general by (3)–(6) thus reduces to

$$[f_{12}(T_1, T_2)S_0(T_1)]^{\delta_0\delta_1}\left[\int_{t_2 = U - T_1}^{\infty} f_{12}(T_1, t_2)\,\mathrm{d}t_2 S_0(T_1)\right]^{(1-\delta_0)\delta_1}[f_0(T_0)S_1(T_0)]^{\delta_0(1-\delta_1)} \times [S_0(U)S_1(U)]^{(1-\delta_0)(1-\delta_1)}. \quad (9)$$

There are two clinical scenarios in which it is desirable to predict survival. At remission the relevant quantity is overall survival. Since we do not know whether a patient will die with or without antecedent relapse, this is given by $T_D = (T_1 + T_2)I[T_1 < T_0] + T_0 I[T_0 < T_1]$. The survivor function of T_D is

$$S_D(t) = \int_{t_1 = 0}^{\infty}\int_{t_2 = \max\{t - t_1, 0\}}^{\infty} f_{12}(t_1, t_2)S_0(t_1)\,\mathrm{d}t_2\mathrm{d}t_1 + \int_{t}^{\infty} f_0(t_0)S_1(t_0)\,\mathrm{d}t_0. \quad (10)$$

Table II. Possible patient outcomes when $K = 1$

Observed data	δ_0	δ_1
$T_1, T_2, I[T_1 < T_0, T_1 + T_2 < U]$	1	1
$T_1, U, I[T_1 < T_0, T_1 < U < T_1 + T_2]$	0	1
$T_0, I[T_0 < \min\{T_1, U\}]$	1	0
$U, I[U < \min\{T_1, T_0\}]$	0	0

At the time of relapse, the relevant quantity is the residual survivor function

$$S_{T_2|T_1}(t_2|t_1) = \Pr[T_2 > t_2 | T_0 > T_1 = t_1] = \int_{t_2}^{\infty} \frac{f_{12}(t_1, u)}{f_1(t_1)}\,\mathrm{d}u \tag{11}$$

which, in particular, does not depend on S_0 due to the assumed independence of T_0 and T_1. Under the bivariate model (8), the residual survival function becomes $S_{T_2|T_1}(t_2|t_1) = S_2(t_2) - \alpha[1 - 2F_1(t_1)]F_2(t_2)S_2(t_2)$, hence the mean residual survival is

$$E[T_2 | T_1 = t < T_0] = \int_0^{\infty} S_2(s)\,\mathrm{d}s - \alpha[1 - 2F_1(t)] \int_0^{\infty} S_2(s)F_2(s)\,\mathrm{d}s$$

which depends explicitly on the association parameter α. This expectation may not exist, however, depending on $F_{1,2}$. A referee has suggested estimation of median rather than mean residual survival, since the former always exists and is robust. Median $[T_2 | T_1 = t < T_0]$ is given by the solution ξ_2 to the equation

$$S_2(\xi_2)[1 - \alpha(1 - 2F_1(t_1))F_2(\xi_2)] = 1/2. \tag{12}$$

3.2. The Colon Cancer Study

We now apply the model to data from a colon cancer trial. Patients with resected stages B and C colorectal carcinoma were randomized to placebo, adjuvant levamisole (lev) or adjuvant levamisole plus fluorouracil (lev + 5-FU) for one year post-surgery. A major goal of the trial was to evaluate the ability of levamisole, either alone or in combination with 5-FU, to prolong the time to relapse and hence survival, based on immunological effects of levamisole in animal models and earlier human trials. A detailed analysis is given by Moertel *et al.*,[10] and these data also are analysed by Lin[11] to illustrate the marginal approach to multivariate time-to-event data. For purposes of illustration here, we restrict attention to the 315 patients on the placebo arm and 304 patients on the lev + 5-FU (treatment) arm and include as covariates a treatment indicator Z_1 and $Z_2 = \log$ (number of nodes involved). Table III summarizes fits of the parametric competing risks model both with and without the two covariates. We obtained standard deviations by bootstrapping. For each of 200 bootstrap samples, we computed the parameter estimates $\hat{\boldsymbol{\theta}}$, and a number of other functions depending on $\hat{\boldsymbol{\theta}}$, such as median residual survival, and then computed the empirical standard deviations of these values.

Figure 2 provides a graphical comparison of the overall survival curves in the two treatment groups, obtained by substituting the parameter estimates into formula (7) and setting log (number

Table III. Competing risks model parameter estimates for colon cancer data

	Death without relapse, T_0	Relapse, T_1	Residual survival, T_2
Without covariates		$\hat{\alpha} = 0{\cdot}990$ (0·063)	
$\hat{\phi}$	1·196 (0·261)	2·233 (0·290)	1·415 (0·113)
$\hat{\lambda}$	55·626 (33.636)	1·027 (0·148)	1·549 (0·143)
$\hat{c}$	0·000 (23·980)	9·199 (1·598)	0·760 (0·265)
Including covariates		$\hat{\alpha} = 0{\cdot}863$ (0·098)	
$\hat{\phi}$	1·215 (0·246)	2·225 (0·399)	1·467 (0·118)
$\hat{\lambda}$	68·823 (58·000)	3·252 (1·044)	3·399 (0·651)
$\hat{c}$	0·000 (26·455)	7·843 (1·945)	0·719 (0·254)
$\hat{\beta}_1$ (lev + 5-FU)	− 0·079 (0·420)	− 1·131 (0·420)	0·460 (0·188)
$\hat{\beta}_2$ (log number of nodes)	0·296 (0·379)	1·929 (0·341)	0·708 (0·171)

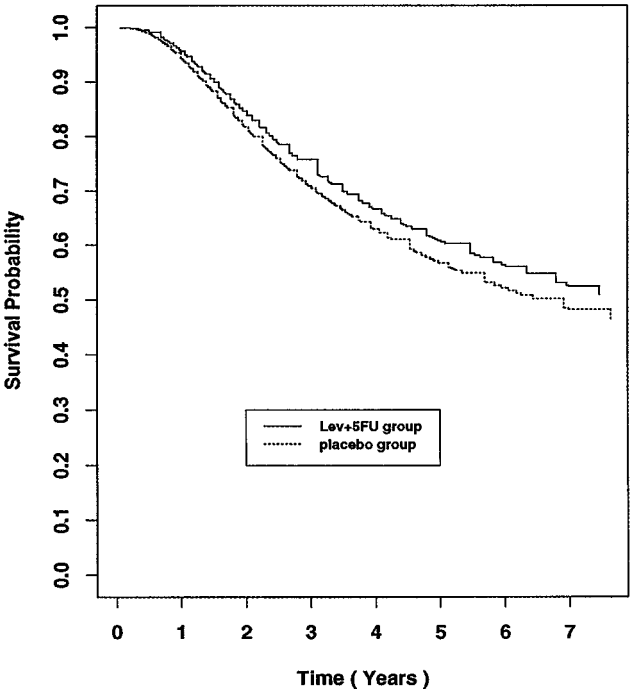

Figure 2. Estimated overall survival curves based on mixture model

of nodes) equal to its mean value 1·33. In terms of overall survival, the superiority of lev + 5-FU over placebo based on our fitted model agrees with the results of Moertel *et al.*[10] and Lin,[11] obtained using different methods.

The fits of the no-covariate model may be assessed graphically by comparing both the estimated overall survival curve and the estimated relapse-free survival curve from the parametric

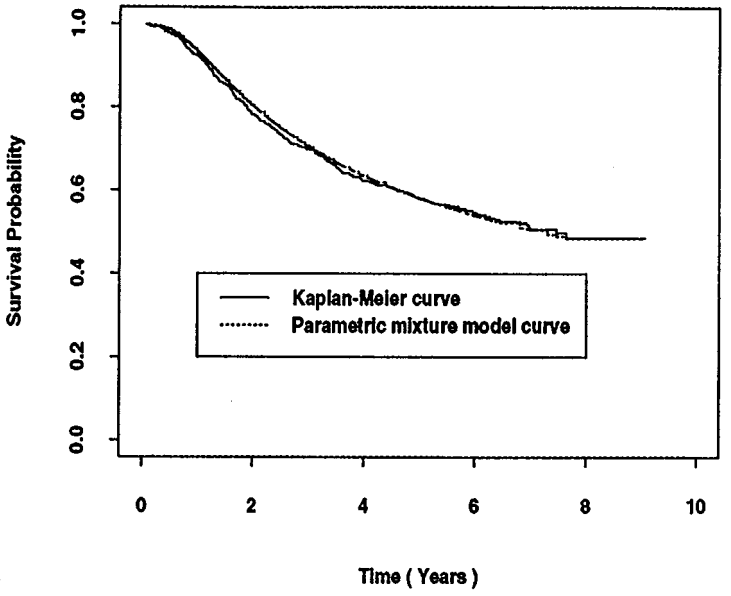

Figure 3. Colon cancer data: overall survival

model to the corresponding Kaplan–Meier curve (Figures 3 and 4). Figures 3 and 4 indicate that the parametric model provides an excellent fit to the data.

Next, we consider the fits of the components of the model, specifically the three marginal distributions and the association parameter. Whereas the treatment effect β_1 on T_0, the time to relapse-free death, is insignificant, the lev + 5-FU group has a significantly longer time to relapse, T_1, but significantly shorter post-relapse survival, T_2. Specifically, $\hat{\beta}_1 = -1{\cdot}131$ for T_1, but $\hat{\beta}_1 = 0{\cdot}460$ for T_2. It thus appears that, among those patients who relapse, the lev + 5-FU group has longer time to relapse but shorter residual survival thereafter. To see how this yields greater overall survival in the lev + 5-FU group, we note that the estimated median of T_1 is 7·4 years in the lev + 5-FU group and 4·6 years in the placebo group, while the corresponding estimated medians of T_2 are 1·2 and 1·7 years, respectively. Thus, since on average T_1 is much larger than T_2 in either treatment group, the superiority of lev + 5-FU over placebo in terms of its effect on T_1 overwhelms its inferiority in terms of the much shorter residual survival T_2.

In terms of $\hat{\alpha}$, the association between T_1 and T_2 is positive and quite strong. That is, on average, patients who take longer to relapse also have longer subsequent residual survival. Note that the degree and sign of association between T_1 and T_2 is a separate issue from the magnitudes and signs of the respective treatment effects on these times.

All three times decreased, on average, with the number of nodes, although this effect was significant only for T_1 and T_2. The addition of the two covariates to the model had little effect on either the $\hat{\phi}_j$'s or $\hat{c}_j$'s, but caused the $\hat{\lambda}_j$'s to increase and $\hat{\alpha}$ to decrease slightly. The fitted models indicate that the distribution of T_0 clearly is Weibull ($c = 0$) and the values $(\hat{\phi}_0 - 1)/s(\hat{\phi}_0) = 0{\cdot}75$ and 0·86 under the two fits indicate further that T_0 follows an exponential distribution. In

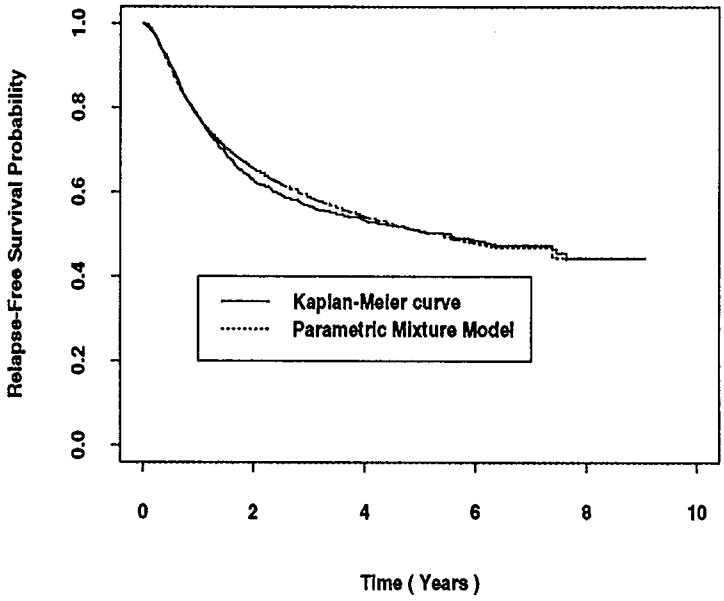

Figure 4. Colon cancer data: relapse-free survival

contrast, the distributions of T_1 and T_2 are neither Weibull nor log-logistic, that is, neither c_1 nor c_2 equals 0 or 1. This illustrates the practical utility of using the three-parameter model for each marginal distribution, and the inadequacy of the two-parameter Weibull or log-logistic distributions. The likelihood ratio statistic for comparing the two models in Table II is 124·11 on 6 d.f., $p < 0{\cdot}0001$, indicating that overall the covariates have a significant effect on the model fit. More specifically, under the second model the 3 d.f. test of $\beta_{0,1} = \beta_{1,1} = \beta_{2,1} = 0$, that is, that there is no overall treatment effect, has LR = 20·8 on 3 d.f., $p = 0{\cdot}00012$. Our result for treatment effect on relapse (T_1) agrees qualitatively with those reported by Lin[11] based on a fit of the marginal model with outcomes T_1 and T_D, in our notation. Since covariate effects under our model correspond to each of T_0, T_1 and T_2 we do not have a single 'treatment effect' parameter corresponding to T_D.

Finally, we consider estimation of median residual life given relapse, as a function of relapse time and the covariates. We computed median residual life given relapse for all observed values of T_1 by substituting the generalized odds rate model (7) of F_1 and F_2 in expression (12), evaluated at the MLE's $\hat{\theta}$, and solving for the median. Figure 5 presents plots of median residual survival given time of relapse, for each of the four patient subgroups determined by the two treatment groups, $Z_1 = 0$ and 1, and the 25th and 75th percentiles of the number of nodes, $Z_2 = \log(2)$ and $\log(5)$. Figure 5 illustrates the negative effects of both lev + 5-FU and the number of nodes on residual survival after relapse. Outcomes for individual patients with given covariates may be predicted in the usual way by computing bootstrap confidence intervals. For example, for placebo patients with 2 nodes the estimated median residual survival given relapse at $t = 2$ years is 1·8 years with 95 per cent bootstrap confidence interval 1·3 to 3·3 years; the corresponding

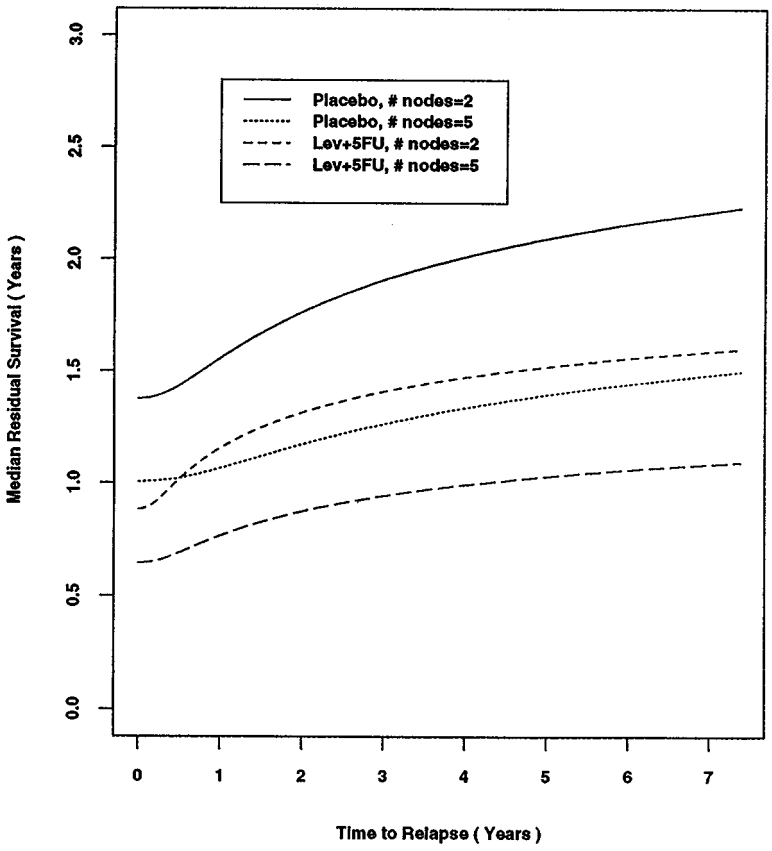

Figure 5. Colon cancer data: median residual survival after relapse

values for lev + 5-FU patients with 2 nodes are 1·2 years with 95 per cent confidence interval 1·0 to 3·1 years.

4. TWO NON-FATAL COMPETING RISKS AND DEATH

4.1. Model

To accommodate two non-fatal competing risks and death, the general model (1) with $K = 2$ is required. In this case there are five marginal distributions corresponding to the times T_0, $T_{1,1}$, $T_{1,2}$, $T_{2,1}$, $T_{2,2}$. Given the marginals, there are two bivariate distributions, $F_{1,12}$ for the time $T_{1,1}$ to the first non-fatal competing risk and subsequent survival time $T_{1,2}$, and likewise $F_{2,12}$ for $(T_{2,1}, T_{2,2})$. Overall survival is

$$T_D = (T_{1,1} + T_{1,2})[T_{1,1} < \min\{T_0, T_{2,1}\}] + (T_{2,1} + T_{2,2})[T_{2,1} < \min\{T_0, T_{1,1}\}] + T_0[T_0 < \min\{T_{1,1}, T_{2,1}\}] \quad (13)$$

Table IV. Possible patient outcomes, death and two non-fatal competing risks

Observed data	δ_0	δ_1	δ_2
$T_{1,1}, T_{1,2}, I[T = T_{1,1} < T_0, T_{1,1} + T_{1,2} < U]$	1	1	0
$T_{1,1}, U, I[T = T_{1,1} < T_0, T_{1,1} < U < T_{1,1} + T_{1,2}]$	0	1	0
$T_{2,1}, T_{2,2}, I[T = T_{2,1} < T_0, T_{2,1} + T_{2,2} < U]$	1	0	1
$T_{2,1}, U, I[T = T_{2,1} < T_0, T_{2,1} < U < T_{2,1} + T_{2,2}]$	0	0	1
$T_0, I[T_0 < \min\{T_{1,1}, T_{2,1}, U\}]$	1	0	0
$U, I[U < \min\{T_{1,1}, T_{2,1}, T_0\}]$	0	0	0

and the overall survivor function is

$$S_D(t) = \int_{t_1=0}^{\infty} \int_{t_2=\max\{t-t_1,0\}}^{\infty} [f_{1,12}(t_1, t_2)S_{2,1}(t_1) + f_{2,12}(t_1, t_2)S_{1,1}(t_1)]S_0(t_1)\mathrm{d}t_2\mathrm{d}t_1 + \int_t^{\infty} f_0(t_0)S_{1,1}(t_0)S_{2,1}(t_0)\mathrm{d}t_0. \quad (14)$$

Due to the assumed independence of T_0, $T_{1,1}$, and $T_{2,1}$ the residual survivor functions are still quite simple, with

$$S_{T_{1,2}|T_{1,1}}(t_2|t_1) = \Pr[T_{1,2} > t_2 | \min\{T_{2,1}, T_0\} > T_{1,1} = t_1] = \int_{t_2}^{\infty} \frac{f_{1,12}(t_1, u)}{f_{1,1}(t_1)}\mathrm{d}u \quad (15)$$

and $S_{T_{2,2}|T_{2,1}}(t_2|t_1)$ defined analogously. Under right censoring there are six possible configurations of patient outcomes, as shown in Table IV. Thus, the model is parameterized by the two dependence parameters α_1, α_2 and $(\lambda_k, \phi_k, c_k, \boldsymbol{\beta}_k)$ for $k = 0$, (1, 1), (1, 2), (2, 1), (2, 2).

4.2. Leukaemia Chemotherapy Data

We illustrate this model by application to a data set provided by E. Estey, M.D. Between 1980 and 1995, 1512 patients with acute myelogenous leukaemia or myelodysplastic syndrome were treated at M.D. Anderson Cancer Center with chemotherapy regimens consisting of cytosine arabinoside alone or in combination with a variety of topo-isomerase II reactive drugs or fludarabine, possibly including either of the growth factors G-CSF or GM-CSF. In chemotherapy of acute leukaemia, the first clinical goal is to bring the patient into complete remission (CR), defined as < 5 per cent circulating blasts, $\geqslant 10^5$ platelets/ml blood, and $\geqslant 10^3$ neutrophils/ml blood. Some patients achieve CR after one course of chemotherapy, some after two courses, and so on. When the physician determines that the patient cannot be brought into CR with a particular chemotherapy regimen, that is, that the patient is *resistant*, the regimen is discontinued in favour of a new treatment, which typically is either a different chemotherapy combination or a bone marrow transplant. Patients who fail to achieve CR are thought to have a substantially decreased chance of long-term survival. Consequently, throughout the course of treatment the physician's primary focus is to achieve and subsequently to maintain CR in the patient. Declaring a patient resistant is essentially a judgement call by the physician, and depends on whether the

Table V. Competing risks model parameter estimates for leukaemia chemotherapy data

	Death without CR or resistance, T_0	Complete remission		Resistance	
		$T_{1,1}$	$T_{1,2}$	$T_{2,1}$	$T_{2,2}$
$\hat{\alpha}$		$-0{\cdot}705$ (0·097)		0·344 (0·174)	
$\hat{\phi}$	1·228 (0·086)	17·991 (1·544)	1·543 (0·099)	3·486 (0·362)	1·124 (0·119)
$\hat{\lambda}$	0·276 (0·041)	0·070 (0·001)	1·062 (0·065)	0·264 (0·018)	0·361 (0·064)
$\hat{c}$	1·798 (0·690)	14·646 (1·500)	2·271 (0·254)	1·321 (0·471)	0·845 (0·336)

physician believes that further treatment with the given regimen is likely to produce a CR. Under our model, T_0 = time to death without CR or resistance, $T_{1,1}$ = time to CR, $T_{1,2}$ = residual survival after CR, $T_{2,1}$ = time to being declared resistant, and $T_{2,2}$ = residual survival after being declared resistant. The times to CR or being declared resistant are non-fatal competing risks, since at most one can occur. Moreover, regimen-related death is a major clinical consideration due to the potentially severe life threatening effects of anti-leukaemia chemotherapy regimens. While intuitively it seems that $T_{1,1}$ and $T_{2,1}$ are correlated, we assume independence due to the general fact that association between competing risks is not identifiable, as noted earlier. More importantly, our proposed methodology applies quite well to the data set considered here since the estimate of the function $\Pr[T_{2,1} < T_{1,1} | T_{1,1} \wedge T_{2,1} > t]$ is an increasing function of t, which reflects the fact that the longer a patient survives without achieving CR the more likely will the physician declare the patient resistant.

To simplify illustration of the competing risks structure, we fit this model to the leukaemia data without including covariates. Table V summarizes the fitted model. Figure 6 gives the fitted overall survival curve and corresponding Kaplan–Meier curve, plotted on the natural log time scale since most of the deaths occur during the first two years. The estimates of the dependence parameter indicate a strong negative association between the time to achieve CR and subsequent residual survival, and a positive association between the time to being declared resistant and residual survival. Figure 7 graphically illustrates these two phenomena. The median residual survival curves given either achievement of CR at time t or given the declaration of resistance at t are plotted as functions of t, along with bootstrap 95 per cent confidence bands. The times of CR are given along the top of the graph, and the times of declaration of resistance along the bottom. The figure shows the precipitous decrease of median residual survival post CR during the first three months of chemotherapy. This illustrates not only the well known fact that CR is the hallmark of survival in chemotherapy of acute leukaemia, but that achievement of CR as quickly as possible during this critical three month period may have a profound effect on subsequent survival.

5. DISCUSSION

We have proposed a class of multivariate parametric models for joint analysis of the times to multiple non-fatal competing risks and death, possibly accounting for covariate effects. The general approach is to model each transition time marginally using a flexible parametric model and then to specify a parsimonious bivariate distribution to characterize the association between the time to each non-fatal event and subsequent survival. We chose the three-parameter

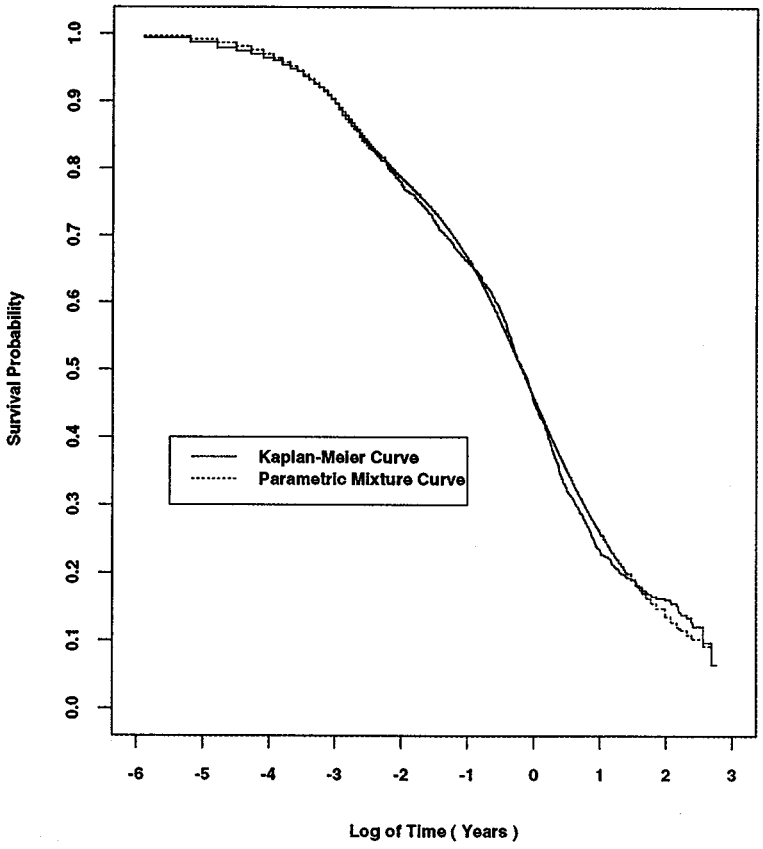

Figure 6. Leukaemia chemotherapy data: overall survival

generalized odds rate distribution for the marginals, due to its flexibility and tractability. We used the von Morgenstern family for each bivariate distribution since it provides a tractable model that characterizes positive or negative association with one parameter. These particular distributions are not essential to our approach in general, however, and one may use any tractable marginal and bivariate distributions with the above properties that are appropriate for a given application. Our formulation avoids the commonly used Markov assumption that, for example, the times to relapse and subsequent survival are independent. In general, the overall survival distribution is a mixture of the survival times corresponding to death without an intervening event or following each possible antecedent non-fatal event. Whereas this requires a more complex model for S_D, our applications suggest that it may be rather unrealistic to ignore the differences between the distributions of T_0 and each $T_{j,1} + T_{j,2}$.

In the case of one non-fatal event, our formulation generalizes Lagakos,[1,2] who used independent exponential distributions for each of T_0, T_1 and T_2. Lagakos was among the first to address the problem of using the observed time of a non-fatal event as an *auxiliary variable* to improve estimation of survival. One may regard the Lagakos model as a special case of the semi-Markov

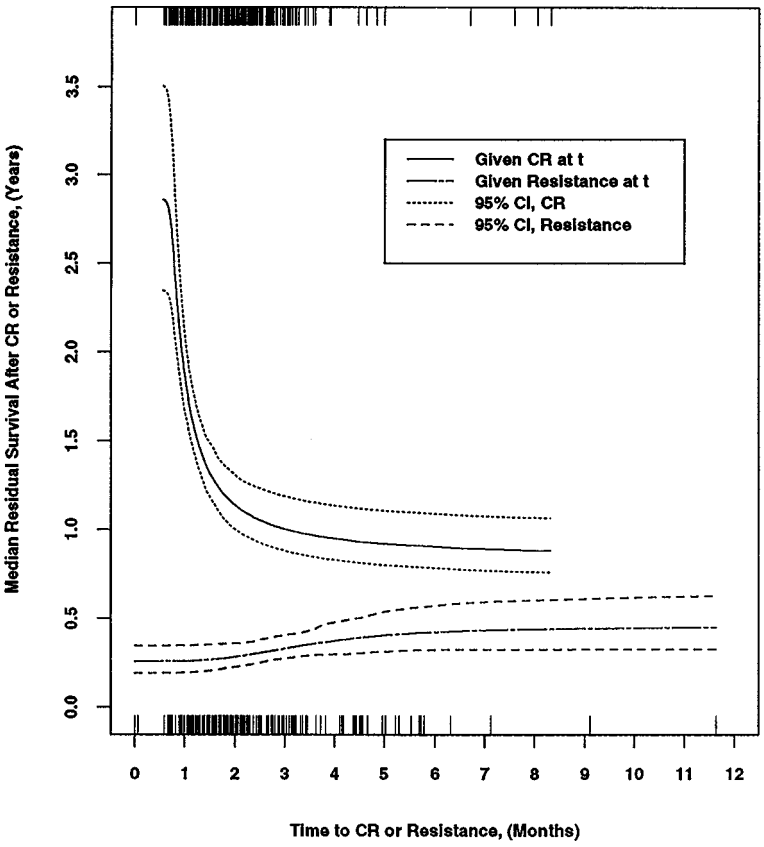

Figure 7. Median residual survival after CR or resistance

process model of Weiss and Zelen[21] or the bivariate exponential model of Freund.[22] Finkelstein and Schoenfeld[3] presented distribution-free methods for assessing overall survival based on the assumption that T_0 and T_1 are independent competing risks and they used a Cox model with T_2 as the outcome variable, either with or without T_1 as a covariate. When T_1 is not included as a covariate, the model assumes that the risk of death among patients who progress does not depend on the time to progression, which also underlies the parametric approach of Lagakos[1] and the non-parametric tests proposed by Hsieh *et al.*[23] Andersen[24] uses a multistate model to analyse the times to diabetic nephropathy (DN) and death with or without DN in insulin-dependent diabetics. His approach comes down to separate Cox regressions for each of the three transition times to DN, death without DN, and death post-DN. As noted by Andersen, however, this does not provide an overall estimate of survival without assumptions similar to those noted above, namely that the death rates for patients with or without DN are the same. Hougaard and Madsen[25] used a similar approach to analyse the times to two non-fatal events and death in heart disease patients who have suffered a myocardial infarction. Gray[5] used kernel smoothing methods to estimate all three marginal distributions while accounting generally for the dependence of T_2 on T_1 and provided both an estimate and a two-sample test for overall survival. Epstein

and Munoz[6] addressed the problem by modelling the overall survival distribution T_D and the conditional distribution of $T_1 I[T_1 \leqslant T_D]$ given T_D, and applied this approach to use presentation time of Kaposi's sarcoma as an auxiliary variable to assess survival in AIDS patients. Kalbfleisch and Lawless[26] presented a pseudo-likelihood approach to the general setting of progression from a healthy state to a diseased state or death. Pepe *et al.*,[27] Pepe[28] and Pepe and Mori[29] provided non-parametric estimators of marginal and conditional probabilities of non-fatal events based on combinations of Kaplan–Meier estimators.

The estimating equation methods of Wei *et al.*[30] and others are aimed primarily at multiple dependent events within a single patient or event times of different patients or individuals which are dependent due to membership in a cluster, such as a litter or family. In illustrating this method, Lin[11] analysed the times to relapse and death in the colon cancer data by fitting Cox models with outcomes T_1 and T_D, hence made no distinction between T_0 and $T_1 + T_2$. When some patients die without relapse, Lin[11] suggested either analysing relapse-free survival time, or, if the proportion of patients who die without relapse is small, assuming independent censoring by death. Our approach provides treatment effects specific to each of T_0, T_1 and T_2. The main advantage of the full likelihood approach over partial likelihood-based estimating equation methods, however, is that a fitted likelihood provides probability estimates. This in turn provides estimates of conditional distributions, medians etc. based on the usual probability calculus, and also lends itself to Bayesian analysis. Another important distinction between the two approaches is that the full likelihood models association parametrically, whereas the estimating equation method is founded on fitting marginally and then accounting for association via a robust estimate of covariance. One may regard this distinction as either an advantage or disadvantage of the likelihood approach.

One alternative to the von Morgenstern distribution for characterizing association between each pair $(T_{j,1}, T_{j,2})$ is a frailty model,[31–34] given generally by $\Pr[T_{j,1} > t_1, T_{j,2} > t_2] = E[\exp(\xi\{\log S_{j,1}(t_1) + \log S_{j,2}(t_2)\})]$, where the frailty ξ is an unobserved non-negative-valued random variable associated with the patient. Although this is reasonably tractable for Weibull marginals and a variety of distributions on ξ, it is much less tractable under log odds rate marginals and allows only positive association. Anderson and Louis[35] have proposed frailty models with negative association, although the tractability of their models in the present context appears limited.

A setting not accommodated by our regime is that where multiple non-fatal events of different types may occur, that is, where they are not competing risks. An obvious but impractical generalization of our approach is to model all transition times marginally, model the joint distribution of each sequence of transition times using a multivariate von Morgenstern distribution, and express T_D as the mixture of the times to death corresponding to all possible paths a patient may take. This leads to a rather unwieldy model, even when $K = 2$. Thus, for such general multivariate time-to-event outcomes, the marginal approach with robust covariance estimate is still the method of choice for most applications. If one desires a full likelihood in such general settings, one needs a parsimonious model that accounts for the important second-order associations.

ACKNOWLEDGEMENTS

We thank Terry Therneau for his thoughtful and constructive comments on an earlier draft of this manuscript, and two referees for their detailed comments and suggestions, which led to

improvements in the manuscript. Thanks also to Eli Estey and Tom Fleming for providing the leukaemia chemotherapy, and the colon cancer data, respectively.

REFERENCES

1. Lagakos, S. W. 'A stochastic model for censored-survival data in the presence of auxiliary variable', *Biometrics*, **32**, 551–559 (1976).
2. Lagakos, S. W. 'Using auxiliary variables for improved estimates of survival time', *Biometrics*, **33**, 399–404 (1977).
3. Finkelstein, D. M. and Schoenfeld, D. A. 'Analyzing survival in the presence of an auxiliary variable', *Statistics in Medicine*, **8**, 1747–1754 (1994).
4. Fleming, T. R., Prentice, R. L., Pepe, M. S. and Glidden, D. 'Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research', *Statistics in Medicine*, **13**, 955–968 (1994).
5. Gray, R. 'A kernel method for incorporating information on disease progression in the analysis of survival', *Biometrika*, **81**, 527–539 (1994).
6. Epstein, L. D. and Munoz, A. 'A bivariate parametric model for survival and intermediate event times', *Statistics in Medicine*, **15**, 1171–1185 (1996).
7. Dobrowska, D. M. and Doksum, K. A. 'Estimation and testing in a two-sample generalized odds-rate model', *Journal of the American Statistical Association*, **83**, 744–749 (1988).
8. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
9. Morgenstern, D. 'Einfache Beispiele zweidimensionaler Verteilungen', *Mittelangsblatt fur Math. Statistik*, **8**, 234–235 (1956).
10. Moertel, C. G., Fleming, T. R., MacDonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., Veeder, M. H. and Mailliard, J. A. 'Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma', *New England Journal of Medicine*, **322**, 352–358 (1990).
11. Lin, D. Y. 'Cox regression analysis of multivariate failure time data: The marginal approach', *Statistics in Medicine*, **13**, 2233–2247 (1994).
12. Gail, M. H. 'A review and critique of some models used in competing risk analysis', *Biometrics*, **31**, 209–222 (1975).
13. Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. 'The analysis of failure time data in the presence of competing risks', *Biometrics*, **34**, 541–554 (1978).
14. Butkiewicz, J. and Hys, E. 'On a class of bi- and multivariate distributions generated by marginal weibull distributions', *Conference on Information Theory, Statistical Decision Functions, Random Processes. A.: Transactions, Prague*, **7**, 73–82 (1977).
15. Johnson, N. L. and Kotz, S. *Distributions in Statistics*: *Continuous Multivariate Distributions*, Wiley, 1972.
16. Bennett, S. 'Log-logistic regression models for survival data', *Applied Statistics*, **32**, 165–171 (1983).
17. Clayton, D. and Cuzick, J. 'The semi-parametric Pareto model for regression analysis of survival times', *Bulletin of the International Statistical Institute* (*Volume 51*). *Proceedings of the 45th Session*, **23.3**, 1–18 (1985).
18. Wingo, D. R. 'Maximum likelihood methods for fitting the Burr type XII distribution to life test data', *Biometrical Journal of Mathematical Methods in Biosciences*, **25**, 77–84 (1983).
19. Gay, D. M. 'A trust region approach to linearly constrained optimization', *in* Lootsma, F. A. (ed.), *Numerical Analysis, Proceedings*, Springer, Berlin, 1983, pp. 171–189.
20. Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap*, Chapman & Hall, 1993.
21. Weiss, G. H. and Zelen, M. 'A semi-Markov model for clinical trials', *Journal of Applied Probability*, **2**, 269–285 (1965).
22. Freund, J. E. 'A bivariate extension of the exponential distribution', *Journal of the American Statistical Association*, **56**, 971–977 (1961).
23. Hsieh, F. Y., Crowley, J. and Tormey, D. C. 'Some test statistics for us in multistate survival analysis', *Biometrika*, **70**, 111–119 (1983).

24. Andersen, P. K. 'Multistate models in survival analysis: A study of nephropathy and mortality in diabetes', *Statistics in Medicine*, **88**, 661–670 (1988).
25. Hougaard, P. and Madsen, E. B. 'Dynamic evaluation of short-term prognosis after myocardial infarction', *Statistics in Medicine*, **4**, 29–38 (1985).
26. Kalbfleisch, J. D. and Lawless, J. F. 'Likelihood analysis of multi-state models for disease incidence and mortality', *Statistics in Medicine*, **7**, 149–160 (1988).
27. Pepe, M. S., Longton, G. and Thornquist, M. 'A qualifier Q for the survival function to describe the prevalence of a transient condition', *Statistics in Medicine*, **10**, 413–421 (1991).
28. Pepe, M. S. 'Inference for events with dependent risks in multiple endpoint studies', *Journal of the American Statistical Association*, **86**, 770–778 (1991).
29. Pepe, M. S. and Mori, M. 'Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data?', *Statistics in Medicine*, **12**, 737–751 (1993).
30. Wei, L. J., Lin, D. Y. and Weissfeld, L. 'Regression analysis of multivariate incomplete failure time data by modeling marginal distributions', *Journal of the American Statistical Association*, **84**, 1065–1073 (1989).
31. Crowder, M. 'A distributional model for repeated failure time measurements', *Journal of the Royal Statistical Society, Series B*, **47**, 447–452 (1985).
32. Clayton, D. and Cuzick, J. 'The EM algorithm for Cox's regression model using GLIM', *Applied Statistics*, **34**, 148–156 (1985).
33. Hougaard, P. 'A class of multivariate failure time distributions', *Biometrika*, **73**, 671–678 (1986).
34. Oakes, D. 'Bivariate survival models induced by frailties', *Journal of the American Statistical Association*, **84**, 487–493 (1989).
35. Anderson, J. and Louis, T. 'Producing negative associations in random effect survival models' (unpublished manuscript).

