J. R. Statist. Soc. B (2003)
65, *Part* 4, *pp.* 921–935

Nonparametric estimation of the sojourn time distributions for a multipath model

Weijing Wang

National Chiao-Tung University, Hsin-Chu, Taiwan

[Received May 2000. Final revision April 2003]

Summary. We use a multipath (multistate) model to describe data with multiple end points. Statistical inference based on the intermediate end point is challenging because of the problems of nonidentifiability and dependent censoring. We study nonparametric estimation for the path probability and the sojourn time distributions between the states. The methodology proposed can be applied to analyse cure models which account for the competing risk of death. Asymptotic properties of the estimators proposed are derived. Simulation shows that the methods proposed have good finite sample performance. The methodology is applied to two data sets.

Keywords: Competing risks; Cure models; Dependent censoring; Identifiability; Illness–death models; Long-term survivors; Multiple end points; Semicompeting risks; Susceptibility

1. Introduction

1.1. Background

Many biomedical applications involve the analysis of multiple end points. Consider a study of bone marrow transplants for leukaemia patients. Some patients will experience recurrence of the malignancy before death but others may die without relapse. A second example is related to the study of a life-threatening acute condition (Betensky and Schoenfeld, 2001). Only a proportion of patients will leave the hospital alive and later die from other causes, whereas the remaining patients, who are not cured, will die in the hospital. These phenomena can be described by the multipath or multistate model that is depicted in Fig. 1. The terminal event, death, leads to an absorbing state whereas the non-terminal event, indicating some progression status, may be bypassed. Let (X, Y, C) be the times to an intermediate end point, a terminal end point and an external censoring end point respectively. Censoring may be due to termination of the study period or the loss of patients to follow-up. Let $\{(X_i, Y_i, C_i),\ i = 1, \ldots, n\}$ be identically and independently distributed replications of (X, Y, C). We shall assume that C is independent of both X and Y, whereas X and Y may be correlated. Statistical methods for analysing a terminal end point and a non-terminal end point are very different. Let $S_1(t) = \Pr(X > t)$.

When death is of primary interest, the Kaplan–Meier estimator is the standard method for estimating the survival function of Y. Progression information can also be incorporated to improve estimation (Gray, 1994). When the prediction of individual survival times is of interest, we need to specify how the intermediate end point affects subsequent survival. Popular choices of model include Markov and semi-Markov models and the proportional hazards model, in which the progression information is treated as a time-dependent covariate. Andersen

Address for correspondence: Weijing Wang, Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan, Republic of China.
E-mail: wjwang@stat.nctu.edu.tw

 1369–7412/03/65921

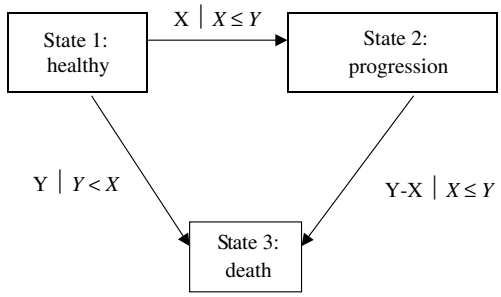

Fig. 1. Multipath model

et al. (1993) and Hougaard (2000) contain many examples and good summaries of related results.

When interest focuses on the progression of disease, statistical issues become more complicated and sometimes controversial (Prentice *et al.*, 1978). Death precludes the biological development of progression whereas censoring terminates observation of the other two events. The major challenge comes from the fact that no progression information is available after death. For those who do not experience progression during their lifetime, X is not well defined and therefore its distribution is non-identifiable without further assumptions. Furthermore, a possible dependence between X and Y complicates a statistical analysis. One popular approach treats death as a dependent competing risk for disease progression. Most methods in the context of competing risks require an explicit specification of the dependent censoring structure; see Zheng and Klein (1995). However, these assumptions are not testable. Furthermore, the implicit assumption underlying many models of the existence of a hypothetical progression distribution after death is debatable. Rather than relying on artificial assumptions, an alternative is to consider only nonparametrically estimable functions, such as the descriptive measures that were proposed by Pepe and Mori (1993). These quantities, however, may have no direct interpretation for the problem of interest.

Another branch of research, under the context of 'long-term survivors' or 'cure models', assumes that only a proportion of subjects will experience the event of interest. Maller and Zhou (1996) provide a useful reference on the subject. This concept may be applied to model the distribution of progression. Let B be an indicator of susceptibility for progression, with $B = 1$ representing susceptibility and $B = 0$ representing immunity or cure. Assuming that $X = \infty$ if $B = 0$, $\Pr(X > t)$ can be written as the mixture form

$$S_1(t) = \Pr(X > t|B = 1)\, p_b + 1 - p_b, \tag{1}$$

where $p_b = \Pr(B = 1)$. When $p_b < 1$, X has an improper distribution. In developing inference procedures for cure models, the definition of cure plays a crucial role. For most methods in the literature, cure is not explicitly defined. Such a model is imposed when a fraction of observations do not experience the event of interest despite long-term follow-up (Li *et al.*, 2001). The majority of references on this rely on the assumption of independent censoring, which does not account for the effect of death. For example, Maller and Zhou (1992) considered estimating $S_1(t)$ by the Kaplan–Meier estimator. They claimed that, when the follow-up time is sufficiently long to observe all the susceptible individuals, the tail probability of the Kaplan–Meier estimator can be used to estimate $1 - p_b$ and hence $\Pr(X > t|B = 1)$ can be estimated via equation (1). Maller and Zhou (1994) further proposed a test for sufficient follow-up on the basis of the distance between the largest and the largest uncensored observations. Although their analysis

looks theoretically appealing, practitioners feel concerned about identifiability of the cure fraction based on nonparametric estimation of the tail probability. Alternatively regression models, which utilize covariate information, have been studied by Farewell (1982), Kuk and Chen (1992) and Taylor (1995), among others. Identifiability of these models has been examined by Li *et al.* (2001). When cures become observable, identifiability is no longer an issue for nonparametric analysis. In Laska and Meisner (1992), cure is observed if a subject does not develop the event of interest within a prespecified time span. Betensky and Schoenfeld (2001) proposed a cure model with random cure times. Using the second example, they defined hospital discharge as the cure event and treated death as a competing risk. The major limitation of their method is that the death and cure times are assumed to be independent of each other.

1.2. Semicompeting risks data

In this paper we consider data of the form $\{(\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y), i = 1, \ldots, n\}$, where $\tilde{X}_i = X_i \wedge Y_i \wedge C_i$, $\tilde{Y}_i = Y_i \wedge C_i$, $\delta_i^x = I(X_i \leqslant Y_i \wedge C_i)$ and $\delta_i^y = I(Y_i \leqslant C_i)$. Fine *et al.* (2001) called such a structure 'semicompeting risks data' since death is a competing risk for progression but not vice versa. Semiparametric estimation procedures for studying the dependence between X and Y when $X \leqslant Y$ have been proposed by Day *et al.* (1997), Fine *et al.* (2001) and Wang (2003). Two-sample comparisons on the basis of X have been studied by Lin *et al.* (1996) and Chang (2000). Without imposing additional assumptions, this problem is non-identifiable owing to dependent censoring by Y. Lin *et al.* (1996) assumed that the distributions of X for the two groups, on a logarithmic scale, follow a bivariate location–shift model, and Chang (2000) assumed a bivariate accelerated failure time model. Although neither specified the relationship between X and Y, they implicitly assumed that the dependence structures are the same for the two groups.

1.3. Main ideas

We consider nonparametric estimation of the following descriptive measures for the multipath model. Let $S_{ij}(t)$ be the survival function of the sojourn time between state i and state j for $i, j = 1, 2, 3$ and $i < j$. Specifically $S_{12}(t) = \Pr(X > t | X \leqslant Y)$, $S_{13}(t) = \Pr(Y > t | X > Y)$ and $S_{23}(t) = \Pr(Y - X > t | X \leqslant Y)$. Also define $S_{123}(t) = \Pr(Y > t | X \leqslant Y)$, which is the survival function for those who will develop progression within their lifespan. Now we discuss applications of these quantities.

Reconsider the problem of a two-sample comparison on X. Let Z be the group indicator, with $Z = 1$ representing the treatment group and $Z = 0$ representing the control group. Treatment could affect both the intensity of X and that of Y from state 1. The difference between $\Pr(X \leqslant Y | Z = 1)$ and $\Pr(X \leqslant Y | Z = 0)$ reveals the treatment effect on the incidence of progression under the competing risk of death. Treatment could also affect transitions between state 1 and state 2 and/or between state 1 and state 3 (as well as between states 2 and 3). The differences of $\Pr(X > t | X \leqslant Y, Z = j)$ for $j = 0, 1$ reflect how the treatment changes the latency distribution of the time to progression if it does occur eventually. These effects may have different scientific meanings and hence should be studied separately.

Because progression cannot be observed after death, it is natural to assign $X = \infty$ if $X > Y$. Letting $\Delta = I(X \leqslant Y)$, we can write

$$S_1(t) = \Pr(X > t | \Delta = 1) p + 1 - p, \tag{2}$$

where $p = \Pr(\Delta = 1)$. Note that expressions (1) and (2) cannot be distinguished on the basis of the information on observed variables, $(\tilde{X}, \tilde{Y}, \delta^x, \delta^y)$, if the cure event B is not observable. It should be noted that the susceptibility model (1) is not sensible if there is any reasonable chance

of a competing risk of $Y < X$. In the second example, where discharge from hospital represents the cure event, it is reasonable to set $\Delta = B$. Therefore the approach that is proposed here can be directly compared with that proposed by Betensky and Schoenfeld (2001). Our method has the advantage that the times to death and cure can be correlated. For the first example of leukaemia relapse, it is questionable to assign $\Delta = B$. If someone dies of an unrelated cause just after entering state 1, it does not make sense to assume that he or she has been cured, or that this person has no chance of progression. Therefore we may want to exclude those who die early from the cured population even though they may not experience recurrence. In such a case, we may define the cure event as $1 - B = I(\Delta = 0, Y > \xi)$, where ξ is a prespecified constant, and the cured rate becomes $(1 - p)\, S_{13}(\xi)$.

2. Methodology

We propose nonparametric estimators for the path probability and the duration distributions between the states. If the path indicator $\Delta = I(X \leqslant Y)$ is known for each subject, nonparametric estimators of these quantities can be easily constructed. When censoring is present, if $\delta^x = 1$, $\Delta = I(X \leqslant Y) = 1$; and, if $\delta^x = 0$ and $\delta^y = 1$, $\Delta = 0$. The major difficulty of estimation comes from double-censored observations with $\delta^x = \delta^y = 0$, for which the value of Δ is unknown, but $X > C$ and $Y > C$, and C is observable. For these observations, we define the conditional path probability given the observed value of C:

$$p(c) = \Pr(X \leqslant Y | X > C, Y > C, C = c) = \frac{1}{H(c)} \int_{v>c} \Pr(X \in [v, v + \mathrm{d}v), Y \geqslant v), \qquad (3)$$

where $H(t) = \Pr(X > t, Y > t)$ is the survival function of $X \wedge Y = \min(X, Y)$. We shall propose an estimator of $p(c)$ and then use it in estimating p, $S_{12}(t)$ and $S_{13}(t)$.

2.1. Estimating an unknown path probability

The objective is to express $p(c)$ as a function of estimable quantities. To simplify the derivation, we assume that Y and C are continuous random variables and X is continuous for $X \leqslant Y$. Our first important result is the identity

$$p(c) = \frac{1}{H(c)} \left\{ \int_{v>c} \frac{\Pr(\tilde{X} \in [v, v + \mathrm{d}v), \tilde{Y} \geqslant v, \delta^x = 1)}{G(v)} + \Pr(\tilde{T} < X \leqslant Y) \right\}, \qquad (4)$$

where $G(v) = \Pr(C > v)$ and $\tilde{T} = \sup\{t : \Pr(X > t, Y > t) \Pr(C > t) > 0\}$. Note that $[0, \tilde{T}]$ is the support of $\tilde{X}$. All the components of $p(c)$ are estimable nonparametrically except the last term in the braces. Specifically the survival function of $X \wedge Y$, $H(t)$, can be estimated by the Kaplan–Meier estimator

$$\hat{H}(t) = \prod_{v \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\tilde{X}_i = v, \tilde{\delta}_i = 1)}{\sum_{i=1}^{n} I(\tilde{X}_i \geqslant v)} \right\},$$

where $\tilde{\delta} = I(X \wedge Y \leqslant C) = \delta^x + \delta^y - \delta^x \delta^y$. There are two versions of the Kaplan–Meier estimator for estimating $G(t)$. Denote $\hat{G}_1(t)$ as the estimator based on $\{(\tilde{X}_i, 1 - \tilde{\delta}_i),\ i = 1, \ldots, n\}$ and $\hat{G}_2(t)$ as the estimator based on $\{(\tilde{Y}_i, 1 - \delta_i^y),\ i = 1, \ldots, n\}$. Now $\Pr(\tilde{X} \leqslant u, \tilde{Y} \geqslant v, \delta^x = 1)$ can be estimated by the empirical function

$$\sum_{i=1}^{n} I(\tilde{X}_i \leqslant u, \tilde{Y}_i \geqslant v, \delta_i^x = 1)/n.$$

We can write $\Pr(\tilde{T} < X \leqslant Y) = p(\tilde{T})\, H(\tilde{T})$. The quantity $H(\tilde{T})$ can be estimated by $\hat{H}(\tilde{X}_{(n)})$, where $\tilde{X}_{(n)}$ is the largest value of $\tilde{X}_i$ $(i = 1, \ldots, n)$. When the support of $X \wedge Y$ lies within $[0, \tilde{T}]$, $H(\tilde{T}) = 0$ and $\Pr(\tilde{T} < X \leqslant Y) = 0$. When the support lies outside $[0, \tilde{T}]$, $p(\tilde{T})$ is not identifiable and additional assumptions are required. Assuming that $p(\tilde{T}) = p$, we can derive an explicit estimator for p whose formula will be given later. For now, let $\tilde{p}(\tilde{X}_{(n)})$ be an estimate of $p(\tilde{T})$. The estimator proposed for $p(c)$ is

$$\hat{p}(c) = \frac{1}{\hat{H}(c)} \left\{ \int_{v>c} \sum_{i=1}^{n} \frac{I(\tilde{X}_i = v, \tilde{Y}_i \geqslant v, \delta_i^x = 1)}{n\, \hat{G}(v)} + \tilde{p}(\tilde{X}_{(n)})\, \hat{H}(\tilde{X}_{(n)}) \right\}, \tag{5}$$

where $\hat{G}(v)$ can be either one of $\hat{G}_j(t)$ $(j = 1, 2)$.

Similar arguments can be applied to the function $q(c) = 1 - p(c)$. Specifically we obtain

$$q(c) = \frac{1}{H(c)} \left\{ \int_{c<v} \frac{\Pr(\tilde{X} \geqslant v, \tilde{Y} \in [v, v + \mathrm{d}v), \delta^x = 0, \delta^y = 1)}{G(v)} + q(\tilde{T})\, H(\tilde{T}) \right\} \tag{6}$$

and

$$\hat{q}(c) = \frac{1}{\hat{H}(c)} \left\{ \int_{v>c} \sum_{i=1}^{n} \frac{I(\tilde{X}_i \geqslant v, \tilde{Y}_i = v, \delta_i^x = 0, \delta_i^y = 1)}{n\hat{G}(v)} + \tilde{q}(\tilde{X}_{(n)})\, \hat{H}(\tilde{X}_{(n)}) \right\}, \tag{7}$$

where $\tilde{q}(\tilde{X}_{(n)}) = 1 - \tilde{p}(\tilde{X}_{(n)})$. In computing $\hat{p}(c)$ and $\hat{q}(c)$, we follow the convention that $0/0 = 0$. Theoretically we only need to estimate either $p(c)$ or $q(c)$ since they sum to 1. For finite samples, the sum is usually not equal to 1. In most of our simulations, $\hat{p}(c) + \hat{q}(c) \approx 1$. In formulae (4) and (6), each observation is weighted by 'the inverse probability of censoring', namely $1/G(v)$, to adjust for the censoring bias. Similar ideas have been used in other estimation problems. Note that, as the value of v increases, $\hat{G}(v)$ approaches closer to 0 and becomes more variable. We shall see in simulations that the performance of $\hat{p}(c)$ and $\hat{q}(c)$ grows worse as the value of c increases. Although $\hat{G}_2(t)$ is a better estimator of $G(t)$ than $\hat{G}_1(t)$ is because $\Pr(C \leqslant Y) \geqslant \Pr(C \leqslant X \wedge Y)$, it is not guaranteed that using $\hat{G}_2(t)$ would produce more efficient estimators of $\hat{p}(c)$ and $\hat{q}(c)$.

In deriving asymptotic properties of all the estimators proposed, we assume that the support of $X \wedge Y$ lies within $[0, \tilde{T}]$, which is the condition of sufficient follow-up. In Appendix A, we prove that, for each $c \in [0, \tilde{T}]$, $\sup_{0 \leqslant c \leqslant \tilde{T}} |\hat{p}(c) - p(c)| \to 0$ with probability 1 and $n^{1/2}\{\hat{p}(c) - p(c)\}$ converges to a mean 0 normal random variable. Properties of $\hat{q}(c)$ can be established by using similar arguments. The asymptotic variance of $\hat{p}(c)$ can be estimated by the moment-type estimator (17), but this is quite complicated and its validity also depends on the condition of sufficient follow-up. Therefore we recommend the bootstrap estimator (18) for variance estimation.

2.2. Estimation of the sojourn time distributions

Now we derive the relationships between $p(c)$ and other quantities of interest. By elementary properties of conditional expectation, it follows that

$$\begin{aligned} \Pr(X \leqslant Y) &= \Pr(\delta^x = 1) + \Pr(X \leqslant Y, \delta^x = \delta^y = 0) \\ &= \Pr(\delta^x = 1) + \int p(c)\, \mathrm{d}\bar{G}_A(c), \\ \Pr(X > Y) &= \Pr(\delta^x = 0, \delta^y = 1) + \int q(c)\, \mathrm{d}\bar{G}_A(c), \end{aligned}$$

where $\bar{G}_A(c) = \Pr(C \leqslant c, X > C, Y > C)$. We can write

$$S_{12}(t) = \prod_{u \leqslant t} \left\{ 1 - \frac{\Pr(\tilde{X} \in [u, u + \mathrm{d}u), \delta^x = 1)}{\Pr(\tilde{X} \geqslant u, X \leqslant Y)} \right\},$$

$$S_{13}(t) = \prod_{v \leqslant t} \left\{ 1 - \frac{\Pr(\tilde{X} \geqslant v, \tilde{Y} \in [v, v + \mathrm{d}v), \delta^x = 0, \delta^y = 1)}{\Pr(\tilde{Y} \geqslant v, X > Y)} \right\},$$

where

$$\Pr(\tilde{X} \geqslant u, X \leqslant Y) = \Pr(\tilde{X} \geqslant u, \delta^x = 1) + \int_{c \geqslant u} p(c)\, \mathrm{d}\bar{G}_A(c),$$

$$\Pr(\tilde{Y} \geqslant v, X > Y) = \Pr(\tilde{Y} \geqslant v, \delta^x = 0, \delta^y = 1) + \int_{c \geqslant v} q(c)\, \mathrm{d}\bar{G}_A(c).$$

The quantity $\bar{G}_A(c)$ can be estimated by $\bar{G}_{A_n}(c) = \Sigma_{i=1}^n I(\tilde{X}_i \leqslant c, \delta_i^x = \delta_i^y = 0)/n$. Using the plug-in principle, we obtain the following nonparametric estimators:

$$\hat{p} = \frac{1}{n} \left\{ \sum_{i=1}^n I(\delta_i^x = 1) + \sum_{i=1}^n I(\delta_i^x = \delta_i^y = 0)\ \hat{p}(\tilde{X}_i) \right\},$$

$$\hat{q} = \frac{1}{n} \left\{ \sum_{i=1}^n I(\delta_i^x = 0, \delta_i^y = 1) + \sum_{i=1}^n I(\delta_i^x = \delta_i^y = 0)\, \hat{q}(\tilde{X}_i) \right\}.$$

Letting $\tilde{p}(\tilde{X}_{(n)}) = \hat{p}$ and $\breve{q}(\tilde{x}_{(n)}) = \hat{q}$, explicit estimators of p and q are given by

$$\hat{p} = \breve{p}(\tilde{X}_{(n)}) = \frac{\sum_{i=1}^n I(\delta_i^x = 1) + \sum_{i \in A_n} \hat{L}_1(\tilde{X}_i)/\hat{H}(\tilde{X}_i)}{n - N_A \hat{H}(\tilde{X}_{(n)})}, \tag{8}$$

$$\hat{q} = \breve{q}(\tilde{x}_{(n)}) = \frac{\sum_{i=1}^n I(\delta_i^x = 0, \delta_i^y = 1) + \sum_{i \in A_n} \hat{L}_2(\tilde{x}_i)/\hat{H}(\tilde{x}_i)}{n - N_A \hat{H}(\tilde{x}_{(n)})}, \tag{9}$$

where $A_n = \{i : \delta_i^x = \delta_i^y = 0\}$,

$$N_A = \sum_{i \in A_n} 1/\hat{H}(\tilde{x}_i),$$

$$\hat{L}_1(c) = \int_{v > c} \sum_{j=1}^n I(\tilde{X}_j = v, \tilde{Y}_j \geqslant v, \delta_i^x = 1)/n\, \hat{G}(v),$$

$$\hat{L}_2(c) = \int_{v > c} \sum_{j=1}^n I(\tilde{X}_j \geqslant v, \delta_j^x = 0, \tilde{Y}_j = v, \delta_j^y = 1)/n\, \hat{G}(v).$$

The sojourn time survival functions can be estimated as

$$\hat{S}_{12}(t) = \prod_{u \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_i^x = 1)}{\sum_{i=1}^n I(\tilde{X}_i \geqslant u, \delta_i^x = 1) + \sum_{i=1}^n I(\tilde{X}_i \geqslant u, \delta_i^x = \delta_i^y = 0)\ \hat{p}(\tilde{X}_i)} \right\}, \tag{10}$$

$$\hat{S}_{13}(t) = \prod_{v \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\tilde{Y}_i = v, \delta_i^x = 0, \delta_i^y = 1)}{\sum_{i=1}^{n} I(\tilde{Y}_i \geqslant v, \delta_i^x = 0, \delta_i^y = 1) + \sum_{i=1}^{n} I(\tilde{Y}_i \geqslant v, \delta_i^x = \delta_i^y = 0)\hat{q}(\tilde{X}_i)} \right\}, \tag{11}$$

$$\hat{S}_{123}(t) = \prod_{v \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\tilde{Y}_i = v, \delta_i^x = 1, \delta_i^y = 1)}{\sum_{i=1}^{n} I(\tilde{Y}_i \geqslant v, \delta_i^x = 1) + \sum_{i=1}^{n} I(\tilde{Y}_i \geqslant v, \delta_i^x = \delta_i^y = 0)\ \hat{p}(\tilde{Y}_i)} \right\}. \tag{12}$$

Asymptotic properties of these estimators also depend on the condition of sufficient follow-up. Consistency and asymptotic normality of $\hat{p}$ and $\hat{S}_{12}(t)$ are proved in Appendices B and C. Similar arguments can be applied to derive the properties of $\hat{q}$, $\hat{S}_{13}(t)$ and $\hat{S}_{123}(t)$. We also suggest use of the bootstrap for variance estimation.

Nonparametric estimation of $S_{23}(t)$ is a different problem since it does not involve the missing path information. Only subjects who have experienced state 2 (i.e. $\delta^x = 1$) will contain useful information for estimating the survival function of $Y - X$. The challenge comes from the problem of dependent censoring. Specifically, for those with $\delta^x = 1$, the larger the value of X, the higher the chance that $Y - X$ will be censored. Using the evolution of acquired immune deficiency syndrome as an example, Wang and Wells (1998) and Lin *et al.* (1999) applied the weighting technique mentioned earlier to handle the effect of dependent censoring. Both methods can be directly applied to estimate $S_{23}(t)$.

2.3. Estimation under an independent censoring model

Many methods in the literature on cure models consider data of the form $\{(\breve{X}_i, \breve{\delta}_i^x), i = 1, \ldots, n\}$, where $\breve{X}_i = X_i \wedge C_i$ and $\breve{\delta}_i^x = I(X_i \leqslant C_i)$. This data structure ignores the competing risk of death, which is inevitable, however. Confusion arises when death without progression happens before the end of the study, i.e. $Y \leqslant C \wedge X$. When this happens, neither $\breve{X}$ nor $\breve{\delta}^x$ is identifiable since X and C are not observable after death. Nevertheless, if the only form of censoring is due to the end of the study, the value of C may be known even if $Y < C$. Further assuming that $X = \infty$ for $X > Y$, we can set $\breve{X} = C$ and $\breve{\delta}^x = 0$. Then $S_1(t)$ can be estimated by the Kaplan–Meier estimator,

$$\breve{S}_1(t) = \prod_{u \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\breve{X}_i = u, \breve{\delta}_i^x = 1)}{\sum_{i=1}^{n} I(\breve{X}_i \geqslant u)} \right\}, \tag{13}$$

and p and $S_{12}(t)$ can be estimated by $\breve{p} = 1 - \breve{S}_1(\breve{X}_{(n)})$ and $\breve{S}_{12}(t) = \{\breve{S}_1(t) - 1 + \breve{p}\}/\breve{p}$ respectively, where $\breve{X}_{(n)}$ is the largest observed value of $\breve{X}_i$ $(i = 1, \ldots, n)$. We shall compare these estimators with the estimators proposed for $S_{12}(t)$ and p via simulations and data analysis. Sometimes practitioners estimate $S_1(t)$ by

$$\tilde{S}_1(t) = \prod_{u \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\tilde{X}_i = u, \delta_i^x = 1)}{\sum_{i=1}^{n} I(\tilde{X}_i \geqslant u)} \right\}. \tag{14}$$

Now we discuss these two estimators of $S_1(t)$. In equation (8), the product–limit estimator is based on estimating the hazard for X^*, where $X^* = X$ provided that $X \leqslant Y$, given that censoring has not occurred. If $X \leqslant Y$ implies that $X < \infty$, $\check{S}_1(t)$ is a valid estimator of $S_1(t)$ for $t \in [0, \tilde{T}]$. Beyond this range, $\check{S}_1(t)$ will overestimate $S_1(t)$. In equation (9), the product–limit estimator is based on estimating the hazard for X^* given that neither death nor censoring has occurred. It can be shown that $\tilde{S}_1(t) \leqslant \check{S}_1(t)$ and $\tilde{S}_1(t)$ will underestimate $S_1(t)$.

3. Numerical results

3.1. Simulation

A series of Monte Carlo simulations was conducted to examine the finite sample performance of the estimators proposed. The algorithm in Prentice and Cai (1992) was used to simulate identically and independently distributed replicates $\{(X_i^0, Y_i),\ i = 1, \ldots, n\}$ from the Clayton model:

$$\Pr(X^0 > x, Y > y) = \{S_1^*(x)^{1-\theta} + S_2(y)^{1-\theta} - 1\}^{1/(1-\theta)}, \qquad \theta > 1,$$

where $S_1^*(x) = \Pr(X^0 > x) = \exp(-\beta x)$ and $S_2(y) = \Pr(Y > y) = \exp(-y)$. The parameter θ controls the degree of association between X^0 and Y and is related to Kendall's τ, $\tau = (\theta - 1)/(\theta + 1)$. To construct the progression time for susceptible individuals, let $X = I(X^0 \leqslant Y)X^0 + I(X^0 > Y)M$, where M is a very large number such that $\Pr(Y > M) \approx 0$. The value of β controls the level of $\Pr(X \leqslant Y) = p$. Censoring variables were generated from $U(0, 6)$. Two types of data, namely semicompeting risks data and the data of independent censoring, were constructed. The censoring proportions are controlled by the values of τ and β. For example, as $\tau = 0.5$ and $\beta = 1$, $\Pr(\delta^x = 0) \approx 60\%$, $\Pr(\delta^y = 0) \approx 18\%$ and $\Pr(\delta^x = \delta^y = 0) \approx 15\%$.

In Table 1, we evaluated the finite sample performances of the estimators proposed for $p(c)$ at some preselected values of c for $n = 100$ and 200. In simulations that are not presented here, we found that the choice of $\hat{G}(t)$ has only a little effect on the resulting estimators. The estimator $\hat{p}(c)$ using $\hat{G}_1(t)$ tends to be a little less variable than that using $\hat{G}_2$. In Table 2, we compare several nonparametric estimators of p and q. Define $\hat{p}_j$ and $\hat{q}_j$ as the estimators proposed for p and q using $\hat{G}(t) = \hat{G}_j(t)$ $(j = 1, 2)$ respectively. We also evaluate the naïve estimator

$$\tilde{p} = \frac{\sum_{i=1}^{n} I(\delta_i^x = 1)}{\sum_{i=1}^{n} \{I(\delta_i^x = 1) + I(\delta_i^x = 0, \delta_i^y = 1)\}},$$

the 'empirical' estimator of p, $\bar{p} = \Sigma_{i=1}^{n} I(\Delta_i = 1)$ and $\check{p} = \check{S}_1(\check{X}_{(n)})$. The difference between the estimators proposed for p and $\bar{p}$ reflects the effect of additional estimation of $\hat{p}(c)$ for double-censored observations. Table 2 shows that $\hat{p}_1$ is a little more efficient than $\hat{p}_2$ but the performances of $\hat{q}_1$ and $\hat{q}_2$ are close. The naïve estimator $\tilde{p}$ appears to be very unstable. It is quite good when p is close to 0.5 but can be very unreliable if p is far from 0.5. In all the cases, $\check{p}$ has a larger variation than the estimators proposed.

In Table 3, we present the summary statistics for the estimators of $S_{12}(t)$ and $S_{13}(t)$ at preselected grid points for $n = 200$. For comparison, we also evaluated the estimators that use the true value of Δ_i for all $i = 1, \ldots, n$. Specifically define

$$\bar{S}_{12}(t) = \prod_{u \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\tilde{X}_i = u, \delta_i^x = 1)}{\sum_{i=1}^{n} I(\tilde{X}_i \geqslant u, \Delta_i = 1)} \right\}, \tag{15}$$

Table 1. Summary statistics of $\hat{p}(c)$ for $(X,Y) \sim \text{Clayton}(\tau = 0.5)$ with $S_1(x) = \exp(-\beta x)$ and $S_2(y) = \exp(-y)$†

c	$p(c)$	*Results for $\beta = 5/6$ and the following values of n:*		$p(c)$	*Results for $\beta = 1$ and the following values of n:*		$p(c)$	*Results for $\beta = 5/4$ and the following values of n:*	
		$n = 100$	$n = 200$		$n = 100$	$n = 200$		$n = 100$	$n = 200$
0.06	0.38	3.1	1.6	0.50	3.7	−0.8	0.64	0.8	0.7
		(5.9)	(4.0)		(5.9)	(4.3)		(5.8)	(4.2)
0.11	0.38	3.9	1.2	0.50	3.4	−0.5	0.65	0.5	0.3
		(6.1)	(4.2)		(6.2)	(4.5)		(6.3)	(4.4)
0.17	0.37	5.4	1.5	0.50	3.1	−0.6	0.66	0.5	0.7
		(6.4)	(4.4)		(6.6)	(4.8)		(6.7)	(4.6)
0.22	0.37	5.2	0.9	0.50	3.3	−0.9	0.66	1.0	−0.4
		(6.7)	(4.6)		(7.0)	(5.0)		(7.0)	(4.9)
0.28	0.36	5.1	1.3	0.50	2.8	−1.0	0.67	0.4	−0.0
		(6.9)	(4.8)		(7.3)	(5.3)		(7.4)	(5.1)
0.33	0.36	5.7	1.0	0.50	2.9	−0.0	0.68	1.3	−0.0
		(7.2)	(5.0)		(7.7)	(5.5)		(7.7)	(5.4)
0.40	0.35	7.1	1.7	0.50	3.0	−0.7	0.69	1.3	−0.2
		(7.5)	(5.2)		(8.1)	(5.7)		(8.3)	(5.7)
0.44	0.35	7.8	2.6	0.50	3.7	−0.6	0.69	0.7	−0.6
		(7.9)	(5.5)		(8.4)	(6.0)		(8.7)	(6.0)
0.50	0.34	8.5	2.9	0.50	2.5	−0.9	0.70	0.9	−0.9
		(8.2)	(5.7)		(9.0)	(6.2)		(8.9)	(6.2)

†In each cell, the first number (times 10^{-3}) is the average bias and the number in parentheses (times 10^{-2}) is the standard deviation of the estimator.

Table 2. Summary statistics for various nonparametric estimators of p and q with $n = 200$†

τ	β	p	$\bar{p}$	$\hat{p}_1$	$\hat{p}_2$	$\hat{q}_1$	$\hat{q}_2$	$\tilde{p}$	$\breve{p}$
0	5/6	0.46	0.30	1.41	1.19	−1.73	−1.73	1.56	1.78
			(3.61)	(3.74)	(3.74)	(3.74)	(3.75)	(3.70)	(3.88)
0	1	0.5	−2.29	−2.28	−2.77	2.32	1.91	−2.36	−3.0
			(3.64)	(3.72)	(3.73)	(3.73)	(3.73)	(3.71)	(3.73)
0	5/4	0.56	−1.66	0.00	−0.61	0.04	−0.49	−0.19	0.02
			(3.43)	(3.58)	(3.58)	(3.58)	(3.56)	(3.55)	(3.58)
0.5	5/6	0.39	−0.31	1.81	1.78	−1.74	−1.96	12.8	−0.18
			(3.56)	(3.82)	(3.82)	(3.83)	(3.86)	(3.83)	(3.93)
0.5	1	0.50	0.13	−0.60	−0.47	0.66	0.57	−0.45	−1.19
			(3.51)	(3.81)	(3.86)	(3.80)	(3.81)	(3.78)	(4.04)
0.5	5/4	0.63	0.72	−0.26	−0.22	0.32	0.17	−9.79	2.07
			(3.57)	(3.59)	(3.65)	(3.59)	(3.60)	(3.60)	(3.94)
0.8	5/6	0.25	−0.88	1.34	1.34	−1.21	−1.18	24.5	−0.25
			(2.93)	(3.16)	(3.16)	(3.16)	(3.17)	(3.37)	(3.17)
0.8	1	0.5	2.0	2.95	2.86	−2.88	−2.98	2.17	2.38
			(3.64)	(4.04)	(4.06)	(4.04)	(4.03)	(3.84)	(4.28)
0.8	5/4	0.78	0.49	−0.18	−0.57	0.24	0.23	−18.6	0.41
			(2.91)	(3.05)	(3.16)	(3.05)	(3.05)	(3.27)	(3.70)

†In each cell, the number (times 10^{-3}) is the average bias and the number in parentheses (times 10^{-2}) is the standard deviation of the estimator based on 500 simulation runs.

Table 3. Summary statistics for the estimators of sojourn time survival functions with $n = 200$, $\tau = 0.5$ and $\beta = 1$†

t	$S_{12}(t)$ and $S_{13}(t)$	$\bar{S}_{12}(t)$	$\hat{S}_{12}(t)$	$\check{S}_{12}(t)$	$\bar{S}_{13}(t)$	$\hat{S}_{13}(t)$
0.105	0.825	−2.4	−2.7	−3.2	−0.5	−0.5
		(4.0)	(4.0)	(4.1)	(3.6)	(3.7)
0.223	0.686	−2.8	−3.5	−4.3	−1.2	−1.1
		(4.6)	(4.7)	(4.8)	(4.5)	(4.6)
0.357	0.570	−2.1	−3.1	−4.3	−1.0	−0.8
		(5.1)	(5.3)	(5.4)	(4.9)	(4.9)
0.511	0.469	1.6	0.4	−1.2	−1.5	−1.2
		(5.1)	(5.4)	(5.6)	(5.0)	(5.1)
0.693	0.378	0.3	−1.0	−3.1	−0.9	−0.6
		(5.0)	(5.4)	(5.7)	(4.9)	(5.1)
0.916	0.295	2.9	1.5	−0.6	−0.3	−0.2
		(4.9)	(5.3)	(5.6)	(4.7)	(5.1)
1.204	0.217	1.5	0.4	−1.9	0.3	0.2
		(4.3)	(4.8)	(5.2)	(4.3)	(4.7)
1.609	0.143	−1.1	−2.0	−4.2	1.3	1.1
		(3.8)	(4.2)	(4.7)	(3.8)	(4.1)
2.302	0.071	−0.6	−1.0	−2.9	0.6	0.0
		(3.0)	(3.4)	(4.0)	(3.0)	(3.2)

†In each cell, the first number (times 10^{-3}) is the average bias and the number in parentheses (times 10^{-2}) is the standard deviation of the corresponding estimator.

$$\bar{S}_{13}(t) = \prod_{v \leqslant t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(\tilde{Y}_i = v, \delta_i^x = 0, \delta_i^y = 1)}{\sum_{i=1}^{n} I(\tilde{Y}_i \geqslant v, \Delta_i = 0)} \right\}. \tag{16}$$

$\check{S}_{12}(t)$ is also more variable than $\hat{S}_{12}(t)$. We also ran simulations under different levels of $\Pr(\delta^x = \delta^y = 0)$. In general the variance of the estimators proposed increases as $\Pr(\delta^x = \delta^y = 0)$ increases.

3.2. Real data examples

The first data set was obtained from Table D.1 of Klein and Moeschberger (1997). There were 137 leukaemia patients receiving bone marrow transplants. Among these patients, 81 died with relapse of leukaemia, only two died without relapse and the remaining 54 patients were doubly censored. We define state 1, *transplantation*, state 2, *relapse*, and state 3, *death*. The naïve estimator of p gives $\tilde{p} = 81/83 = 0.976$. Because $\hat{H}(\tilde{X}_{(n)}) = \hat{H}(2640) = 0.336 > 0$, we applied formulae (8) and (9) and then obtained $\hat{p} = 0.963$ and $\hat{q} = \check{q}(2640) = 0.023$ using $\hat{G}(t) = \hat{G}_1(t)$. Because $\hat{p} \approx 1$, this data set may not be a typical example of the multipath model that is considered here. It turns out that $\hat{S}_1(t)$ and $\hat{S}_{12}(t)$ are very close. Since there were only two patients with $\delta^x = 0$ and $\delta^y = 1$, $\tilde{S}_1(t)$ and $\hat{S}_1(t)$ are very close.

The second data set is the Stanford heart transplantation data (Crowley and Hu, 1977). Define state 1, *acceptance*, state 2, *transplantation*, and state 3, *death*. Among 103 participants, 69 received transplants, 30 died without transplantation and only four observations were double censored. The naïve estimator gives $\tilde{p} = 69/99 = 0.697$. The data set provided the date of acceptance, date of transplant, date last seen and an indicator for the status of dead or alive.

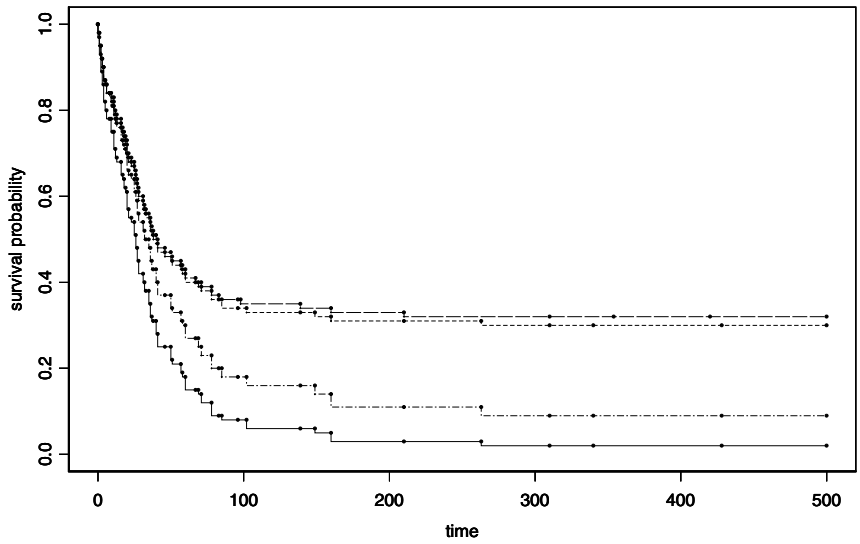

Fig. 2. Estimated survival probability of transplantation time based on the heart transplant data: – – –, $\check{S}_1(t)$; -------, $\hat{S}_1(t)$; · · · · · ·, $\tilde{S}_1(t)$; ———, $\hat{S}_{12}(t)$

We computed $\hat{H}(\tilde{X}_{(n)}) = \hat{H}(1400) = 0.02$, which implies that the missing path proportion is tiny. There were two cases of $\hat{p}(c)$ whose values exceed 1 but the values of $\hat{q}(c)$ looked more reasonable. Therefore we use formula (7) for $\hat{q}(c)$ and set $\hat{p}(c) = 1 - \hat{q}(c)$. For the four observations with $\delta^x = \delta^y = 0$, the estimated path probabilities are $\hat{q}(1401) = 0.304$, $\hat{q}(428) = 0.304$, $\hat{q}(31) = 0.386$ and $\hat{q}(11) = 0.361$ which give $\hat{p} = 0.71$ and $\hat{q} = 0.304$. Note that there were 30 patients with $\delta^x = 0$ and $\delta^y = 1$. For these observations, we set $\check{X} = C$, which is the difference between the date of acceptance and the end of the study in April 1974, and $\check{\delta}^x = 0$. The four estimated curves of $\hat{S}_{12}(t)$, $\hat{S}_1(t)$, $\tilde{S}_1(t)$ and $\check{S}_1(t)$ are plotted in Fig. 2. We found that $\check{S}_1(t)$ has a plateau at $\check{q} = 0.318$, which is close to $\hat{q} = 0.304$, and hence $\hat{S}_1(t)$ and $\check{S}_1(t)$ are close. It turns out that the incorrect estimator $\tilde{S}_1(t)$ has a much lighter tail than $\check{S}_1(t)$ and hence is misleading. In Fig. 3, we present the three estimated curves, $\hat{S}_{13}(t)$, $\hat{S}_2^{\mathrm{KM}}(t)$ and $\hat{S}_{123}(t)$, where $\hat{S}_2^{\mathrm{KM}}(t)$ denotes the Kaplan–Meier estimator of $\Pr(Y > t)$. It is well known in the medical literature that the phenomenon $\hat{S}_{123}(t) > \hat{S}_{13}(t)$ might be attributed to selection bias instead of the transplant effect.

4. Concluding remarks

We used the framework of a multipath model to describe data with multiple end points. The methodology proposed was applied to cure models. It is well known that, when no cures are observable, the cured fraction is not identifiable (Li *et al.*, 2001). Our analysis shows that the problem of non-identifiability is partly due to the competing risk of death. An important fact is that the commonly used data structure, $\{(\check{X}_i, \check{\delta}_i^x),\ i = 1, \ldots, n\}$, is by itself not identifiable if $\Pr(\delta^x = 0, \delta^y = 1) > 0$. Therefore the multipath model, which accounts for the competing risk of death, is a more natural formulation for studying the cure population. In the literature of cure models, 'cure' does not have a single definition. We used two examples to illustrate possible relationships between Δ and B. When $\Delta = B$, both the method proposed and the Kaplan–Meier estimator can estimate the cured fraction and the latency distribution for the susceptible individuals. Our approach has several advantages over the Kaplan–Meier estimator. Although both meth-

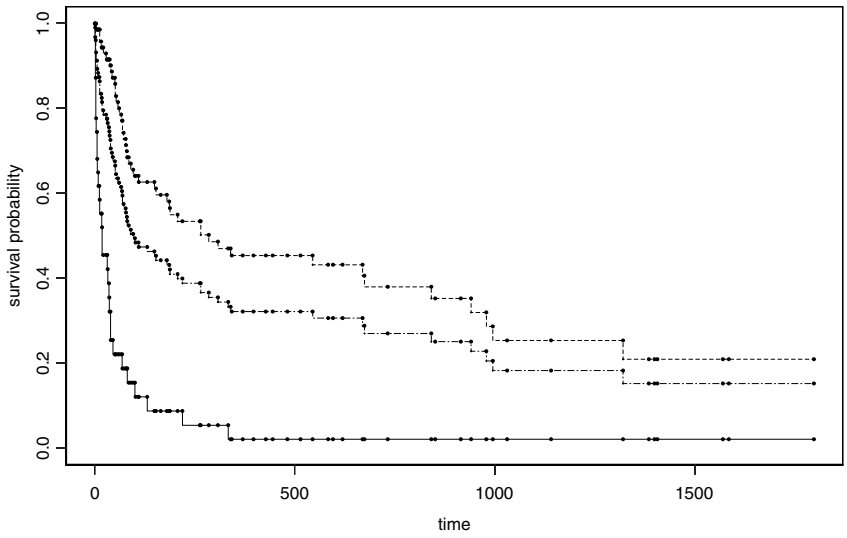

Fig. 3. Estimated survival probability of death time based on the heart transplant data: -------, $\hat{S}_{123}(t)$; · · · · · ·, $\hat{S}_2^{KM}(t)$; ———, $\hat{S}_{13}(t)$

ods rely on the condition of sufficient follow-up, our approach is more robust if this assumption is violated. Specifically, under our setting, as long as $H(\tilde{T})$ is close to 0 or the assumed value of $p(\tilde{T})$ is reasonable, the bias of $\hat{p}(c)$ is still minimal. The method proposed allows the dependence between death and progression and hence is more general than the method proposed by Betensky and Schoenfeld (2001). Furthermore our approach also provides estimators of $S_{13}(t)$ and $S_{123}(t)$, which may be useful in other contexts. For example when $1 - B = I(\Delta = 0, Y > \xi)$ indicates the cure status, $S_{13}(\xi)(1 - p)$ measures the cure probability. The technique that was used in estimating p and the sojourn time distributions is similar to the E-step in the EM algorithm. The missing values of the path indicator for double-censored observations are estimated by their conditional expected values, which are then used to estimate sufficient statistics for the quantities of interest.

Acknowledgements

The author is grateful to the Joint Editor, the referees and Dr Chen-Hsin Chen, Dr Tony Chen and Dr Adam Ding for their helpful comments on the work. The author also thanks J. Hsieh for performing the data analysis. The research for this paper was funded by National Science Council grant 86-2115-M-001-033-T.

Appendix A: Asymptotic properties of $\hat{p}(c)$

Under sufficient follow-up, $p(c) = L(c)/H(c)$, where

$$L(c) = \int_{c<u<\tilde{T}} K_u(\mathrm{d}u)/G(u)$$

and

$$K_v(u) = \Pr(\tilde{X} \leqslant u, \tilde{Y} \geqslant v, \delta^x = 1).$$

We can write $\hat{p}(c) = \hat{L}(c)/\hat{H}(c)$, where

$$\hat{L}(c) = \int_{u>c} \hat{K}_u(\mathrm{d}u)/\hat{G}(u)$$

and

$$\hat{K}_v(u) = \sum_{i=1}^{n} I(\tilde{X}_i \leqslant u, \delta_i^x = 1, \tilde{Y}_i \geqslant v)/n.$$

We now prove strong consistency of $\hat{p}(c)$. Since $H(\cdot)$ is non-increasing, $1/H(c) \leqslant 1/H(\tilde{T}) = M_1 < \infty$ for all $0 \leqslant c < \tilde{T}$. Similarly $1/G(c) \leqslant 1/G(\tilde{T}) = M_2 < \infty$ for all $0 \leqslant c < \tilde{T}$. By strong consistency of the Kaplan–Meier estimators and applying the triangle inequality, one can show that, with probability 1,

$$\sup_{0\leqslant c<\tilde{T}} |\hat{p}(c) - p(c)| \leqslant M_1 \sup_{0\leqslant c<\tilde{T}} |\hat{L}(c) - L(c)| + M_1 \sup_{0\leqslant c<\tilde{T}} |\hat{H}(c) - H(c)|.$$

By properties of the limit supremum function, strong consistency of the Kaplan–Meier estimators and applying integration by parts, one can show that, for any $\varepsilon > 0$,

$$\Pr\Big[\limsup\Big\{\omega : \sup_{0\leqslant c\leqslant\tilde{T}} |\hat{p}(c,\omega) - p(c)| \geqslant \varepsilon\Big\}\Big] = 0,$$

where ω is an element in the probability space.

Now we show asymptotic normality of $n^{1/2}\{\hat{p}(c) - p(c)\}$. We can write

$$\frac{1}{p(c)} n^{1/2}\{\hat{p}(c) - p(c)\} \stackrel{\mathrm{a}}{=} \frac{1}{L(c)} n^{1/2}\{\hat{L}(c) - L(c)\} - \frac{1}{H(c)} n^{1/2}\{\hat{H}(c) - H(c)\}.$$

It can be shown that $n^{1/2}\{\hat{L}(c) - L(c)\}/L(c) =^{\mathrm{a}} a_n + b_n$, where

$$a_n = \frac{1}{L(c)}\Big[- \frac{1}{G(c)} n^{1/2}\{\hat{K}_c(c) - K_c(c)\} + \int_{v>c} \frac{n^{1/2}\{\hat{K}_v(v) - K_v(v)\}}{G^2(v)}\, G(\mathrm{d}v)\Big],$$

$$b_n = -\frac{1}{L(c)} \int_{v>c} \frac{n^{1/2}\{\hat{G}(v) - G(v)\}}{G^2(v)}\, K_v(\mathrm{d}v).$$

By weak convergence of $n^{1/2}\{\hat{K}_v(v) - K_c(v)\}$ and $n^{1/2}\{\hat{G}(v) - G(v)\}$ to mean 0 Gaussian processes, asymptotic normality of b_n and hence of $n^{1/2}\{\hat{p}(c) - p(c)\}$ can be established.

Each component of $n^{1/2}\{\hat{p}(c) - p(c)\}$ can be further expressed explicitly as the sum of mean 0 random variables. Specifically we can write $a_n = n^{-1/2}\Sigma_{i=1}^n A(\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y, c)$, where

$$A(\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y, c) = \frac{1}{L(c)}\Big[- \frac{1}{G(c)}\{I(\tilde{X}_i \leqslant c, \tilde{Y}_i \geqslant c, \delta_i^x = 1) - \Pr(\tilde{X} \leqslant c, \tilde{Y} \geqslant c, \delta^x = 1)\}$$
$$+ \int_{v>c} \frac{1}{G^2(v)}\{I(\tilde{X}_i \leqslant v, \delta_i^x = 1, \tilde{Y}_i \geqslant v) - \Pr(\tilde{X} \leqslant v, \tilde{Y} \geqslant v, \delta^x = 1)\}\, \mathrm{d}G(v)\Big].$$

If we use $\hat{G}_2(t) = \hat{G}(t)$, it follows that

$$\frac{1}{G(v)} n^{1/2}\{\hat{G}(v) - G(v)\} \stackrel{\mathrm{a}}{=} -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \int_0^v \frac{M_i^c(\mathrm{d}u)}{\Pr(\tilde{Y} \geqslant u)},$$

where $M_i^c(u) = \int_0^u I(\tilde{Y}_i \leqslant s, \delta_i^y = 0) - \int_{s\leqslant u} I(\tilde{Y}_i \geqslant s)\, \Lambda_c(\mathrm{d}s)$ and $\Lambda_c(\mathrm{d}s) = \Pr(C \in [s, s + \mathrm{d}s) | C \geqslant s)$, and hence $b_n = n^{-1/2}\Sigma_{i=1}^n B(\tilde{Y}_i, \delta_i^y, c)$, where

$$B(\tilde{Y}_i, \delta_i^y, c) = \frac{1}{L(c)} \int_{v>c} \frac{1}{G(v)} \int_{0<u\leqslant v} \frac{\mathrm{d}M_i^c(\mathrm{d}u)}{\Pr(\tilde{Y} \geqslant u)} K_v(\mathrm{d}v).$$

Also

$$\frac{1}{H(c)} n^{1/2}\{\hat{H}(c) - H(c)\} \stackrel{\mathrm{a}}{=} -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \int_0^c \frac{M_i^{xy}(\mathrm{d}u)}{\Pr(\tilde{X} \geqslant u)} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} C(\tilde{X}_i, \tilde{\delta}_i, c),$$

where

$$C(\tilde{X}_i, \tilde{\delta}_i, c) = \int_0^c M_i^{xy}(\mathrm{d}u)/\Pr(\tilde{X} \geqslant u),$$

$$M_i^{xy}(u) = \int_0^u I(\tilde{X}_i \leqslant s, \tilde{\delta}_i = 0) - \int_{s \leqslant u} I(\tilde{X}_i \geqslant s)\, \Lambda_{xy}(\mathrm{d}s)$$

and

$$\Lambda_{xy}(\mathrm{d}s) = \Pr(X \wedge Y \in [s, s + \mathrm{d}s) | X \geqslant s, Y \geqslant s).$$

We have shown that

$$\frac{1}{p(c)} n^{1/2} \{\hat{p}(c) - p(c)\} \stackrel{\mathrm{a}}{=} \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{A(\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y, c) + B(\tilde{Y}_i, \delta_i^y, c) + C(\tilde{X}_i, \tilde{\delta}_i, c)\},$$

which converges in distribution to a mean 0 normal random variable with variance $\Sigma_c = E[\{A(\tilde{X}, \tilde{Y}, \delta^x, \delta^y) + B(\tilde{Y}, \delta^y) + C(\tilde{X}, \tilde{\delta})\}^2]$. Therefore a moment-type estimator of the asymptotic variance of $\hat{p}(c)$ is

$$\hat{p}^2(c) \sum_{i=1}^{n} \{\hat{A}(\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y, c) + \hat{B}(\tilde{Y}_i, \delta_i^y, c) + \hat{C}(\tilde{X}_i, \tilde{\delta}_i, c)\}^2 / n^2, \tag{17}$$

where $\hat{A}$, $\hat{B}$ and $\hat{C}$ are obtained by plugging in nonparametric estimates of the unknown components in A, B and C respectively. The variance estimator (17) depends on the assumption of sufficient follow-up. The bootstrap method provides a more flexible alternative for variance estimation. Let $\{(\tilde{X}_i^*, \tilde{Y}_i^*, \delta_i^{x^*}, \delta_i^{y^*}), i = 1, \ldots, n\}$ be a bootstrapped sample. The resampling procedure is repeated B times. Let $\hat{p}_b^*(c)$ be the estimator proposed for $p(c)$ based on the bth bootstrapped sample. The asymptotic variance of $\hat{p}(c)$ can be estimated by

$$\sum_{b=1}^{B} \{\hat{p}_b^*(c) - \hat{p}(c)\}^2 / (B - 1). \tag{18}$$

Appendix B: Asymptotic properties of $\hat{p}$

It follows that

$$|\hat{p} - p| \leqslant \left| \sum_{i=1}^{n} I(\delta_i^x = 1)/n - \Pr(\delta^x = 1) \right| + \left| \int_c \hat{p}(c)\, \mathrm{d}\bar{G}_{A_n}(c) - \int_c p(c)\, \mathrm{d}\bar{G}_A(c) \right|.$$

By the law of large numbers, the first term converges to 0 in probability. The second term is bounded by $r_{1n} + r_{2n}$, where $r_{1n} = \int_c |\hat{p}(c) - p(c)|\, \mathrm{d}\bar{G}_{A_n}(c)$ and $r_{2n} = |\int_c p(c)\{\mathrm{d}\bar{G}_{A_n}(c) - \mathrm{d}\bar{G}_A(c)\}|$. By uniform and strong consistency of $\hat{p}(c)$ and the bounded convergence theorem, $r_{1n} = o_p(1)$. Applying integration by parts to r_{2n} and by strong consistency of the empirical estimator $\bar{G}_{A_n}(c)$, $r_{2n} = o_p(1)$. It follows that $n^{1/2}(\hat{p} - p) = p_{1n} + p_{2n} + p_{3n} + o_p(1)$, where

$$p_{1n} = n^{-1/2} \sum_{i=1}^{n} \{I(\delta_i^x = 1) - \Pr(\delta^x = 1)\},$$

$$p_{2n} = n^{-1/2} \sum_{i=1}^{n} \int p(c)\{\hat{A}(\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y, c) + \hat{B}(\tilde{Y}_i, \delta_i^y, c) + \hat{C}(\tilde{X}_i, \tilde{\delta}_i, c)\}\, \mathrm{d}G_A(c),$$

$$p_{3n} = \int p(c)\{\mathrm{d}\bar{G}_{A_n}(c) - \mathrm{d}\bar{G}_A(c)\}\sqrt{n}.$$

Asymptotic normality of p_{1n}, p_{2n} and p_{3n} can be established by using standard techniques. The bootstrap method also provides a practical solution for variance estimation.

Appendix C: Asymptotic properties of $\hat{S}_{12}(t)$

We can write $S_{12}(t) = \Pi_{u \leqslant t}\{1 - F_{10}(\mathrm{d}u)/R_{12}(u)\}$, where $F_{10}(u) = \Pr(\tilde{X} \leqslant u, \delta^x = 1)$ and $R_{12}(u) = \Pr(\tilde{X} \geqslant u, X \leqslant Y)$, which can be estimated by

$$\hat{R}_{12}(u) = \frac{1}{n}\sum_{i=1}^{n} I(\tilde{X}_i \geqslant u, \delta_i^x = 1) + \int_{c \leqslant u} \hat{p}(c)\,\mathrm{d}\bar{G}_{A_n}(c)$$

and $\hat{F}_{10}(u) = \Sigma_{i=1}^{n} I(\tilde{X}_i \leqslant u, \delta_i^x = 1)/n$. Asymptotic normality of $\hat{S}_{12}(t)$ can be established by showing that it is a smoothed function of $\hat{R}_{12}(u)$ and $\hat{F}_{10}(u)$, each of which is consistent and asymptotically normal.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models based on Counting Processes*. New York: Springer.

Betensky, R. A. and Schoenfeld, D. A. (2001) Nonparametric estimation in a cure model with random cure times. *Biometrics*, **57**, 282–286.

Chang, S. H. (2000) A two-sample comparison for multiple ordered event data. *Biometrics*, **56**, 183–189.

Crowley, J. and Hu, M. (1977) Covariance analysis of heart transplant survival data. *J. Am. Statist. Ass.*, **72**, 27–36.

Day, R., Bryant, J. and Lefkopoulou, M. (1997) Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika*, **84**, 45–56.

Farewell, V. T. (1982) The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041–1046.

Fine, J. P., Jiang, H. and Chappell, R. (2001) On semi-competing risks data. *Biometrika*, **88**, 907–919.

Gray, R. (1994) A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika*, **81**, 527–539.

Hougaard, P. (2000) *Analysis of Multivariate Survival Data*. New York: Springer.

Klein, J. P. and Moeschberger, M. L. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

Kuk, A. Y. C. and Chen, C. (1992) A mixture model combining logistic regression with proportional hazards regressions. *Biometrika*, **79**, 531–541.

Laska, E. M. and Meisner, M. J. (1992) Nonparametric estimation and testing in a cure model. *Biometrics*, **48**, 1223–1234.

Li, C.-S., Taylor, J. M. G. and Sy, J. P. (2001) Identifiability of cure models. *Statist. Probab. Lett.*, **54**, 389–395.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996) Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*, **83**, 381–393.

Lin, D. Y., Sun, W. and Ying, Z. (1999) Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika*, **86**, 59–70.

Maller, R. A. and Zhou, S. (1992) Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**, 731–739.

Maller, R. A. and Zhou, S. (1994) Testing for sufficient follow-up and outliers in survival data. *J. Am. Statist. Ass.*, **89**, 1499–1506.

Maller, R. A. and Zhou, S. (1996) *Survival Analysis with Long-term Survivors*. New York: Wiley.

Pepe, M. S. and Mori, M. (1993) Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statist. Med.*, **12**, 737–751.

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

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

Taylor, J. M. G. (1995) Semi-parametric estimation in failure time mixture models. *Biometrics*, **51**, 899–907.

Wang, W. (2003) Estimating the association parameter for copula models under dependent censoring. *J. R. Statist. Soc.* B, **65**, 257–273.

Wang, W. and Wells, M. T. (1998) Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, **85**, 561–572.

Zheng, M. and Klein, J. P. (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82**, 127–138.