Published in final edited form as:

Aust N Z J Stat. 2010 ; 52(1): 75–91. doi:10.1111/j.1467-842X.2009.00567.x.

NONPARAMETRIC ESTIMATION OF CONDITIONAL CUMULATIVE HAZARDS FOR MISSING POPULATION MARKS

Dipankar Bandyopadhyay and Amalia Jácome Pumar

Medical University of South Carolina, USA and Universidade da Coruña, Spain

Summary

A new function for the competing risks model, the conditional cumulative hazard function, is introduced, from which the conditional distribution of failure times of individuals failing due to cause j can be studied. The standard Nelson–Aalen estimator is not appropriate in this setting, as population membership (mark) information may be missing for some individuals owing to random right-censoring. We propose the use of imputed population marks for the censored individuals through fractional risk sets. Some asymptotic properties, including uniform strong consistency, are established. We study the practical performance of this estimator through simulation studies and apply it to a real data set for illustration.

Keywords

competing risks; fractional risk set; Nelson–Aalen; right-censoring

1. Introduction

Consider a multistate model in which a healthy (state 0) subject may end up with one of J different types of failure. This model, called ‘competing risk’ in the statistical literature, is relevant in many disciplines, such as medicine, demography, actuarial science (as multiple decrement models), economics and manufacturing; see Crowder (2001) and Lindqvist (2006). For each unit, we observe both the time to failure T^* and the type of failure $X^* \in \{1, \dots, J\}$. In this setup, the observed failure times are frequently right-censored, possibly owing to the termination of the study before all the subjects fail, or because subjects are lost to follow-up. Consequently, T^* and X^* are unknown for the right-censored individuals. To accommodate this situation, we include an additional random variable C , the latent time to censoring, which is observed when a subject is right-censored.

The study of competing risks considers the crude probability $F_j(t) = \Pr(T^* \leq t, X^* = j)$, also known as the cumulative incidence function, and the cause-specific cumulative hazard function

$$\Lambda_j(t) = \int_0^t \frac{dF_j(v)}{1 - F(v^-)}, \quad (1)$$

where $F = \sum_{j=1}^J F_j$ is the cumulative distribution function (cdf) of T^* . Neither function may be appropriate if one desires to understand the summary probability of the different causes of failure, given that failures have already occurred. For our study, we focus on the estimation of the j th conditional functions

$$F_j^*(t) = \Pr(T^* \leq t | X^* = j) \text{ and } \Lambda_j^*(t) = \int_0^t \frac{dF_j^*(v)}{1 - F_j^*(v^-)}. \quad (2)$$

This approach is suitable when the target is the complete distribution and not the subdistribution function of the lifetimes associated with each risk separately as in (1).

For a better explanation of this problem, we consider the following real-life example. A clinical trial of estrogen diethylstilbestrol (DES) was carried out in which patients with stage 3 or 4 prostate cancer were assigned to four treatment groups (Cheng, Fine & Wei 1998; Escarela & Carrière 2003). Patients died of either ‘prostate cancer’ ($j = 1$) or ‘other causes’ ($j = 2$). Also considered were right-censored individuals, whose eventual cause of death was unknown, however the cause of death might be due to any one of the two causes. The

conditional distribution function F_j^* is more appropriate to answer the question: ‘Among all the patients who died of prostate cancer (in a situation where we would know the cause of failure of all the individuals), what is the chance of surviving to age 60?’ whereas, the crude probability F_j is the answer to the specific question: ‘What is the chance that a patient will die of prostate cancer before age 60?’. We revisit this data set in Section 5.

The rest of the paper is organized as follows. In Section 2, we introduce the Nelson–Aalen type estimator of the conditional cumulative hazard function based on fractional risk sets (FRSs), and in Section 3 we give some asymptotic properties, including a strong representation and consistency. In Section 4, we consider a small simulation study to assess the finite sample performance of the novel estimator, and apply it to a real data set in Section 5. We conclude with a discussion in Section 6, followed by an Appendix, which contains the proofs of the main results.

2. Estimation based on fractional risk sets

We consider the competing risk network as a multistate continuous-time stochastic process $\{Z(t), t \in \mathcal{T}\}$ with a finite-state space $S = \{1, \dots, J, 0\}$ having a tree topology and right-continuous sample paths $Z(t^+) = Z(t)$, where we assume that the states $1, \dots, J$ are absorbing whereas state 0 is transient (the root node). Here $\mathcal{T} = [0, \tau]$, where τ is a large, possibly observed, time point ($\leq \infty$). Typically, for applications, τ will be taken to be the largest time at which some event (failure) took place. Let T_i^* be the time at which the i th individual leaves stage 0 for a failure (stage j , say), with cumulative distribution function (cdf) F , and let X_i^* denote the stage occupied by the i th individual at time T_i^* (i.e., its failure type). In the presence of right-censored observations, we record the censoring time C with cdf G . Hence, although the variables of interest are T^* and X^* , one cannot observe (T_i^*, X_i^*) , but (T_i, δ_i, X_i) , with $T_i = \min(T_i^*, C_i)$, the right-censored failure time with cdf H , the censoring indicator $\delta_i = \mathbf{1}(T_i^* \leq C_i)$, and the observable cause of failure $X_i = X_i^* \delta_i$. If an individual is not censored, then $X_i^* = X_i$; otherwise, $X_i = 0$ and X^* is unknown. It is further assumed that the censoring variable C is independent of (T^*, X^*) , and that all the random variables are independent and identically distributed (i.i.d.) across the n individuals.

There are no data-driven estimators of Λ_j^* and F_j^* in the literature. One may be tempted to estimate these functions using classical ideas, as in Andersen *et al.* (1993):

$$\widehat{\Lambda}_j^*(t) = \int_0^t \frac{\mathbf{1}(Y_j(v) > 0)}{Y_j(v)} dN_j(v) = \sum_{T_i \leq t} \frac{\mathbf{1}(X_i = j)}{Y_j(T_i)} \quad \text{and}$$

$$\widehat{F}_j^*(t) = 1 - \prod_{T_i \leq t} \left(1 - \frac{\mathbf{1}(X_i = j)}{Y_j(T_i)} \right),$$

where $N_j(t) = \sum_{i=1}^n \mathbf{1}(T_i \leq t, X_i = j)$ is the number of observed failures of type j in the time interval $[0, t]$, and

$$Y_j(t) = \sum_{i=1}^n \mathbf{1}(T_i \geq t, X_i^* = j) \quad (3)$$

is the number of individuals at risk of failing due to cause j or of being censored. The functions $\widehat{\Lambda}_j^*$ and $\widehat{F}_j^*$ are not estimators themselves, as the size of the subpopulation at-risk set Y_j is not computable with the observations (T_i, δ_i, X_i) . In the absence of such an identifier, we can still assign a probability of each individual being in one of the J subpopulations. We estimate Y_j by Y_j^f , the ‘fractional risk set’ corresponding to the j th cause of failure (Satten & Datta 1999; Datta & Satten 2000), defined as

$$Y_j^f(t) = \sum_{i=1}^n \widehat{\phi}_{ij} \mathbf{1}(T_i \geq t). \quad (4)$$

Here, $\widehat{\phi}_{ij}$ is the estimated probability that the i th individual belongs to the j th subpopulation according to its failure type:

$$\widehat{\phi}_{ij} = \begin{cases} 1, & \text{if } X_i = j \\ 0, & \text{if } X_i > 0, X_i \neq j, \\ \widehat{P}_j(T_i, \infty), & \text{if } X_i = 0 \end{cases}$$

where

$$\widehat{P}_j(s, t) = \int_{(s,t]} \left(\prod_{(s,u]} \left(1 - \frac{dN_j(u)}{Y_j(u)} \right) \right) \frac{dN_j(u)}{Y_j(u)}, \quad (5)$$

with $N(t) = \sum_{j=1}^J N_j(t)$ and $Y(t) = \sum_{i=1}^n \mathbf{1}(T_i \geq t)$, is the Aalen–Johansen estimator (Andersen *et al.* 1993, pp. 288–289), and the nonparametric maximum likelihood estimator (NPMLE) of the transition probability is

$$\begin{aligned} P_j(s, t) &= P(T^* \leq t, X^* = j \mid T^* > s) \\ &= P(\text{Failure type } j \text{ by time } t \mid \text{alive at time } s). \end{aligned} \quad (6)$$

Then, $Y_j^f(t)$ gives the estimated fractional mass in the j th failure group remaining at risk of failure at time t , counting ($\hat{\phi}_{ij} = 1$) the observations that failed due to cause j , discarding ($\hat{\phi}_{ij} = 0$) the observations that failed due to a cause other than j , and estimating the probability of being in the j th failure-type group ($\hat{\phi}_{ij} = \hat{P}_j(T_i, \infty)$) for the censored observations. Thus, the proposed estimators of Λ_j^* and F_j^* are

$$\widehat{\Lambda}_j^{*f}(t) = \sum_{T_i \leq t} \frac{\mathbf{1}(X_i = j)}{Y_j^f(T_i)} \text{ and } \widehat{F}_j^{*f}(t) = 1 - \prod_{T_i \leq t} \left(1 - \frac{\mathbf{1}(X_i = j)}{Y_j^f(T_i)} \right),$$

with Y_j^f given in (4). In fact, the proportion $\pi_j = \Pr(X^* = j)$ of the sample that fails due to cause j can be estimated as

$$\widehat{P}_j(0, \infty) = \frac{1}{n} \sum_{i=1}^n \widehat{\phi}_{ij}, \quad j=1, \dots, J.$$

The ‘fractional risk set’ concept has recently been used in the literature with multistate models; see Satten & Datta (1999), Datta & Satten (2000) and Bandyopadhyay & Datta (2008).

3. Asymptotic properties

As a first step, we provide some results for $\hat{P}_j$ and Y_j^f , which are of independent interest in themselves. They will be used to derive the main properties of $\widehat{\Lambda}_j^{*f}$. The analogous results for $\widehat{F}_j^{*f}$ can be derived considering the one-to-one mapping relation

$$1 - F(t) = \exp(-\Lambda_c(t)) \prod_{u \leq t} (1 - \Delta\Lambda(u)),$$

where Λ_c is the continuous part of Λ , and $\Delta\Lambda(u) = \Lambda(u) - \Lambda(u^-)$. We first obtain, in Theorem 1, an i.i.d. representation of $\hat{P}_j$. This is a novel result that can be used to derive many other asymptotic properties of $\hat{P}_j$, for example the strong consistency result in Corollary 1, which is analogous to the consistency in Aalen (1978, theorem 1) for partial transition probabilities and to that in Fleming (1978, theorem 5.1) for nonhomogeneous Markov processes. The same properties will be obtained, in Theorem 2 and Corollary 2 respectively, for the fractional risk-set estimator Y_j^f .

From these properties for $\hat{P}_j$ and Y_j^f , and the relationship

$$\widehat{\Lambda}_j^{*f}(t) = \widehat{\Lambda}_j^*(t) + \int_0^t \left(\frac{1}{Y_j^f(v)} - \frac{1}{Y_j(v)} \right) dN_j(v), \quad (7)$$

we obtain the strong i.i.d. representation of $\widehat{\Lambda}_j^{*f}$ in Theorem 3, and the uniform strong consistency in Corollary 3.

Before we state the main results, we introduce some notation. Fix $b_H < \sup\{t : H(t) < 1\}$, with H the cdf of the observable lifetime $\mathcal{T}$, and consider the subdistribution functions (c , censored; nc , not-censored)

$$H^{nc}(t) = \Pr(T \leq t, \delta > 0), \quad (8)$$

$$H^c(t) = \Pr(T \leq t, \delta = 0) \text{ and } \bar{H}^c(t) = \Pr(T \geq t, \delta = 0), \quad (9)$$

$$H_j^{nc}(t) = \Pr(T^* \leq t, \delta > 0, X^* = j) = \Pr(T \leq t, X = j), \quad (10)$$

$$H_j(t) = \Pr(T \leq t, X^* = j) \text{ and } \bar{H}_j(t) = \Pr(T \geq t, X^* = j). \quad (11)$$

Consider now the empirical estimators of (10) and (11) given by

$$H_{nj}^{nc}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \leq t, X_i = j) = \frac{1}{n} N_j(t),$$

$$H_{nj}(t^-) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i < t, X_i^* = j) = \frac{1}{n} (n_j - Y_j(t)),$$

where $n_j = \sum_{i=1}^n \mathbf{1}(X_i^* = j)$. The importance of these functions is clear, as the Nelson–Aalen type estimator $\widehat{\Lambda}_j^*(t)$ of the conditional cumulative hazard function can also be written as follows:

$$\widehat{\Lambda}_j^*(t) = \int_0^t \frac{dH_{nj}^{nc}(v)}{n_j/n - H_{nj}(v^-)}.$$

The following theorem gives an i.i.d. representation of the estimator of the transition probabilities $\hat{P}_j$.

Theorem 1. *If the distribution functions F and G are continuous, then*

$$\widehat{P}_j(s, t) - P_j(s, t) = \frac{1}{n} \sum_{i=1}^n \zeta_j(T_i, X_i^*, \delta_i, s, t) + r_n(s, t),$$

where

$$\begin{aligned}
& \zeta_j(T, X^*, \delta, s, t) \\
&= \frac{1}{1 - F(s)} \\
&\times \left(\frac{1 - F(T)}{1 - H(T)} \mathbf{1}(s \leq T \leq t, \delta > 0, X^* = j) \right. \\
&- \int_s^t \frac{1 - F(v)}{(1 - H(v))^2} \mathbf{1}(T \geq v, X^* = j) dH^{\text{nc}}(v) \\
&- \frac{1}{1 - H(T)} \int_s^t (1 - F(v)) \mathbf{1}(s \leq T \leq v, \delta > 0) d\Lambda_j(v) + \int_s^t (1 - F(v)) \int_0^v \frac{\mathbf{1}(T \geq u)}{(1 - H(u))^2} dH^{\text{nc}}(u) d\Lambda_j(v) \Big) \\
&- P_j(s, t) \int_0^s \frac{\mathbf{1}(T \geq v)}{(1 - H(v))^2} dH^{\text{nc}}(v),
\end{aligned} \tag{12}$$

with H^{nc} given in (8) and $\sup_{0 \leq s \leq t \leq b_H} |r_n(s, t)| = O(n^{-1} \ln n)$ almost surely.

Remark 1. This result generalizes the i.i.d. representation for the Aalen–Johansen estimator evaluated at (t, ∞) , which is important in the study of many statistical properties in competing risks:

$$\widehat{P}_j(t, \infty) - P_j(t, \infty) = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - F(t)} \int_t^\infty \frac{1 - F(v)}{1 - H(v)} \times (dM_{ji}(v) - P_j(v, \infty) dM_{.i}(v)) + r_n(t),$$

where

$$M_{ji}(t) = \mathbf{1}(T_i \leq t, \delta_i > 0, X_i^* = j) - \int_0^t \mathbf{1}(T_i \geq s, X_i^* = j) d\Lambda(s), \quad M_{.i}(t) = \sum_{j=1}^J M_{ji}(t)$$

and $\sup_{0 \leq t \leq b_H} |r_n(t)| = O(n^{-1} \ln n)$ almost surely.

Corollary 1. The estimator of the transition probabilities satisfies

$$\sup_{0 \leq s \leq t \leq b_H} n^{1/2} (\ln n)^{-1/2} |\widehat{P}_j(s, t) - P_j(s, t)| \rightarrow 0.$$

Theorem 2. If the distribution functions F and G are continuous, then

$$\frac{1}{n} (Y_j(t) - Y_j^{\text{f}}(t)) = \frac{1}{n} \sum_{i=1}^n \rho_j(T_i, X_i^*, \delta_i, t) + s_n(t),$$

where

$$\rho_j(T, X^*, \delta, t) = \mathbf{1}(T \geq t, \delta = 0) (\mathbf{1}(X^* = j) - P_j(T, \infty)) - \int_t^\infty \zeta_j(T, X^*, \delta, v, \infty) d\overline{H}^{\text{c}}(v), \tag{13}$$

with H^{c} given in (9), ζ_j in (12) and $\sup_{0 \leq t \leq b_H} |s_n(t)| = O(n^{-1} \ln n)$ almost surely.

Corollary 2. *The fractional risk-set estimator Y_j^f satisfies*

$$\sup_{0 \leq t \leq b_H} n^{-1/2} (\ln n)^{-1/2} |Y_j(t) - Y_j^f(t)| \rightarrow 0.$$

The following theorem gives a representation of the FRS estimator of Λ_j^* as a sum of i.i.d. variables plus a remainder term. It is based on the strong representations for the transition probabilities in Theorem 1 and for the FRS in Theorem 2.

Theorem 3. *If the distribution functions F and G are continuous, then*

$$\widehat{\Lambda}_j^{*f}(t) - \Lambda_j^*(t) = \frac{1}{n} \sum_{i=1}^n \xi_j(T_i, X_i^*, \delta_i, t) + R_n(t),$$

where

$$\xi_j(T, X^*, \delta, t) = \frac{\mathbf{1}(T \leq t, X^* = j)}{\bar{H}_j(T)} - \int_0^t \frac{\mathbf{1}(T \leq v, X^* = j)}{\bar{H}_j^2(v)} dH_j^{\text{nc}}(v) + \int_0^t \frac{\rho_j(T, X^*, \delta, v)}{\bar{H}_j^2(v)} dH_j^{\text{nc}}(v)$$

with H_j^{nc} given in (10), $\bar{H}_j$ in (11), ρ_j in (13), and

$$\sup_{0 \leq t \leq b_H} |R_j(t)| = O(n^{-1} (\ln n)^3) \text{ almost surely.}$$

Corollary 3. *The fractional risk-set estimator of Λ_j^* satisfies*

$$\sup_{0 \leq t \leq b_H} \left(\frac{n}{(\ln n)^3} \right)^{1/2} \left| \widehat{\Lambda}_j^{*f}(t) - \Lambda_j^*(t) \right| \rightarrow 0 \text{ almost surely.}$$

Remark 2. The nature of the dependence between T^* (time to failure) and X^* (cause of failure) is very useful. Under independence, both variables can be studied separately, which greatly simplifies the analysis of competing risks. There are few tests in the literature for the independence between T^* and X^* . Without censoring, Dewan, Deshpande & Kulathinal (2004) proposed several tests based on the conditional probabilities $\phi_j(t) = \Pr(X^* = j \mid T^* > t) = P_j(t, \infty)$, as T^* and X^* are independent if and only if ϕ_j is a constant.

With censored observations, Dykstra, Kochar & Robertson (1998) and Kochar & Proschan (1991) provided some restricted tests. More recently, tests of independence using FRSs were studied by Bandyopadhyay & Datta (2008). In this context, the conditional functions F_j^* and Λ_j^* could be used to test independence without any restriction as, in that case, $F_j^* = F$ and $\Lambda_j^* = \Lambda$.

4. Simulation study

We carried out a small simulation study to assess the practical performance of the FRS estimator of Λ_j^* . Note that, when T^* and X^* are independent, we have $\Lambda_j^* = \Lambda$ for $j = 1, \dots, J$. So, we also considered the Nelson–Aalen estimator of Λ ,

$$\widehat{\Lambda}_n^{\text{NA}}(t) = \sum_{T_i \leq t} \frac{\mathbf{1}(\delta_i > 0)}{Y(T_i)} \quad \text{with } Y(T_i) = \sum_{k=1}^n \mathbf{1}(T_k \geq T_i),$$

to estimate Λ_j^* under the independence of T^* and X^* . However, the independence assumption is not always appropriate, and $\widehat{\Lambda}_j^{*\text{f}}$ will be the only available estimator of Λ_j^* .

We simulated two models with $J = 2$ competing risks. For Model 1 (El-Nouty & Lancar 2004), we assume the independence of T^* and X^* , whereas in Model 2 the variables T^* and X^* are dependent. In both models, T^* is the lifetime (in days) and takes values in the interval $[0, \tau]$. The parameter τ was chosen equal to 1461 to represent 4 years. In this case, the percentage of censoring is about 8.33% in Model 1 and 33% in Model 2.

Model 1. The variables T^* and X^* are independent, with

$$\begin{aligned} F(t) &= 1 - (1 - t/\tau)^{3/4} \exp(-t^2/(2\tau^2)) \quad \text{and } G(t) = 1 - (1 - t/\tau)^{1/4} \exp(t^2/(2\tau^2)) \\ F_1(t) &= F_2(t) = 0.5F(t) \\ \Lambda_j^*(t) &= t^2/(2\tau^2) - 3/4 \ln(1 - t/\tau) \quad \text{for } j=1, 2. \end{aligned}$$

Model 2. The variables T^* and X^* are dependent, with

$$\begin{aligned} F(t) &= t/\tau \quad \text{and } G(t) = 1 - (1 - t/\tau)^2 \\ F_1(t) &= t/(2\tau)(1 - (1 - t/\tau)^2) \quad \text{and } F_2(t) = t/(2\tau)(1 + (1 - t/\tau)^2) \\ \Lambda_1^*(t) &= -\ln(1 - t/\tau(1 - (1 - t/\tau)^2)) \quad \text{and } \Lambda_2^*(t) = -\ln(1 - t/\tau(1 + (1 - t/\tau)^2)). \end{aligned}$$

For Model 1, $\varphi_j(t) = P(X^* = j) = 0.5$ for $j = 1, 2$, and therefore the variables T^* and X^* are independent. However, for Model 2, we have

$$\varphi_1(t) = 0.5 \left(1 + \frac{t}{\tau} - \frac{t^2}{\tau^2} \right) \quad \text{and } \varphi_2(t) = 0.5 \left(1 - \frac{t}{\tau} + \frac{t^2}{\tau^2} \right),$$

and, as φ_j is not constant, T^* and X^* are dependent.

Figure 1 shows the FRS estimator $\widehat{\Lambda}_j^{*\text{f}}$ for $j = 1, 2$ and the NA estimator $\widehat{\Lambda}_n^{\text{NA}}$. Under the independence of T^* and X^* (Model 1), we have $\Lambda_1^* = \Lambda_2^* = \Lambda$, and the three estimators are very similar. However, when T^* and X^* are dependent (Model 2), $\widehat{\Lambda}_j^{*\text{f}}$ and $\widehat{\Lambda}_n^{\text{NA}}$ differ from each other. Hence, the plot of the FRS and NA estimators can be used as a preliminary test for the independence of T^* and X^* .

Under independence, the best estimator for Λ_j^* is $\widehat{\Lambda}_n^{\text{NA}}$ rather than $\widehat{\Lambda}_j^{*\text{f}}, j = 1, 2$. In order to show this, we computed the mean squared error (MSE) of the FRS and NA estimators as (see Figure 2)

$$MSE(\widehat{\Lambda}_j^{*\text{f}}(t)) = E((\widehat{\Lambda}_j^{*\text{f}}(t) - \Lambda_j^*(t))^2) \text{ and } MSE(\widehat{\Lambda}_n^{\text{NA}}(t)) = E((\widehat{\Lambda}_n^{\text{NA}}(t) - \Lambda_j^*(t))^2).$$

In Model 1 (T^* and X^* are independent), the NA estimator is preferable for estimating Λ_j^* as it has a smaller MSE. However, under dependence (Model 2), $MSE(\widehat{\Lambda}_n^{\text{NA}})$ is larger than $MSE(\widehat{\Lambda}_j^{*\text{f}})$. Hence, we infer that the NA estimator is not a serious candidate for estimating Λ_j^* unless there is clear evidence for the independence of T^* and X^* , and the FRS estimator should be used in any case.

The FRS-based estimators can also be used to check the independence of T^* and X^* using a Kolmogorov–Smirnov (KS)-type test. Under independence, we have $F_j^* = F, j = 1, \dots, J$.

Thus, considering the FRS estimator of F_j^* and the Kaplan–Meier (KM) estimator of F , we construct the following KS-type test:

$$T_{\text{KS}} = \sup_{t \in \mathbb{R}} \max_{j=1, \dots, J} |\widehat{F}_j^{*\text{f}}(t) - F_n^{\text{KM}}(t)|.$$

Figure 3 provides the histogram of the values of the KS test computed using $m = 1000$ Monte Carlo samples of size $n = 100$. As expected, T_{KS} takes much lower values in Model 1, where T^* and X^* are independent, than in Model 2. This shows that the KS test is promising for testing independence. However, the theoretical large-sample properties of this test, together with size and power evaluations, are outside the scope of this paper and will be addressed in the future.

5. Example: prostate cancer data

In this section, we illustrate the FRS estimators using the prostate cancer data described in Section 1. We consider the same data set as studied by Byar & Green (1980) and published in Andrews & Herzberg (1985). In those papers, the randomized trial was aimed at comparing different levels of DES, a drug used to treat prostate cancer, with respect to patient survival. A total of 506 patients were assigned to four treatment groups. Because of the potentially fatal cardiovascular adverse effect from DES (Escarela & Carrière 2003), the assessment of risk–benefit analysis of DES (Cheng *et al.* 1998) must take into account not only the death time from prostate cancer ($j = 1$), but also other competing causes of death ($j = 2$), including death from cardiovascular-related causes. Although we do not consider the set of covariates, 23 patients with incomplete covariate information were removed from our analysis. Out of the 483 patients, there were 125 (about 26%) deaths from prostate cancer, 219 (about 45%) deaths from ‘other causes’, along with 139 (about 29%) right-censored observations whose subpopulation membership of the eventual cause of failure was unknown, however it was assumed to be one of the two causes for our problem. We are interested in the distribution of the time to death from prostate cancer ($j = 1$) and other different causes ($j = 2$), and also in verifying the dependence/independence between the lifetime and the cause of failure.

First, we estimate the distribution functions F_j^* using the FRS estimator. Figure 4 shows that, among all the patients who died because of prostate cancer (including censored individuals

who may die of prostate cancer), the chance of surviving to month 60 after administration of DES is 0.87. Note that the FRS estimates of F_j^* for $j = 1$ and $j = 2$ are very similar, which suggests that the lifetime and cause of failure are independent.

We also computed the FRS estimator of Λ_j^* for $j = 1$ and 2 and the NA estimator $\widehat{\Lambda}_n^{\text{NA}}$. Figure 5 shows that the FRS estimators are very close to each other, and quite different from $\widehat{\Lambda}_n^{\text{NA}}$. This agrees with the conclusion of the log-rank test defined in Bandyopadhyay & Datta (2008) that tests the null hypothesis $H_0: F_1^*(t) = \dots = F_J^*(t)$, which is equivalent to testing the independence of T^* and X^* . The log-rank test statistic for this data set is 5.989, with a bootstrap estimated standard error of 14.275, leading to a p -value of 0.176, which is not significant at the 5% level. We can conclude that, among the patients who died of (a) prostate cancer and (b) other causes, the application of the drug DES did not provide any differential effect on the survival behaviour of the two competing groups.

6. Discussion

In this article, two new functions, the j th conditional distribution function F_j^* and the corresponding cumulative hazard function Λ_j^* , were introduced for the competing risks model. This helps us to describe the distribution of the lifetimes of those individuals failing from a specific cause j when there is a set of J interacting causes.

A new Nelson–Aalen-type estimator of Λ_j^* was proposed for when the population marks of some individuals are unknown owing to right-censoring. The key is to split the total risk set at every time point t into J possible at-risk subsets, also called the FRS. The FRS methodology provides a way to include the censored observations in the estimation with a nice probability interpretation. Some properties of the FRS estimator $\widehat{\Lambda}_j^{*f}$ have been derived, such as an asymptotic i.i.d. representation and a strong consistency result, together with the same results for the transition probability estimator $\widehat{P}_j$ and the FRS estimator Y_j^f . Our simulation studies show that, in addition to providing information on lifetimes of individuals failing due to a specific cause, FRS estimators of F_j^* and Λ_j^* can be used to test the independence of T^* and X^* .

Although we have restricted our attention to the estimation of the conditional cumulative hazards under a competing risk framework, this can be extended much further to more complicated multistate networks, for example the three-stage irreversible illness–death model (Andersen *et al.* 1993, pp. 28–34). The conditional cumulative hazards among different states in the multistate model can be compared using FRSs. This will be pursued elsewhere.

Acknowledgments

Part of this work was done while the first author was a doctoral student at the University of Georgia, USA. He acknowledges research support from the University of Georgia through a Graduate School Dissertation Completion Fellowship. This research was partially supported by grants P20 RR017696-06 and 5U10 DA013727-09 from the United States National Institutes of Health. The second author would like to acknowledge economic support through grant MTM2008-00166 from the Spanish Ministerio de Ciencia e Innovación. The authors thank Professor Somnath Datta for his valuable comments and for drawing attention to the fractional risk-set methodology. We are also grateful to the Associate Editor and a referee for their critical comments, which led to a substantially better presentation of the manuscript.

Appendix

Proof of Theorem 1. The transition probabilities in (6) can also be written as follows:

$$P_j(s, t) = \frac{1}{1 - F(s)} \int_s^t (1 - F(v)) d\Lambda_j(v),$$

with F the cdf of T^* , and Λ_j the cause-specific cumulative hazard function given in (1). For the Aalen–Johansen estimator of $P_j(s, t)$ in (5), we have

$$\widehat{P}_j(s, t) = \frac{1}{1 - F_n^{\text{KM}}(s)} \int_s^t (1 - F_n^{\text{KM}}(v)) d\widehat{\Lambda}_j^{\text{NA}}(v),$$

where F_n^{KM} is the product-limit Kaplan–Meier (KM) estimator, and $\widehat{\Lambda}_j^{\text{NA}}$ the Nelson–Aalen (NA) estimator of Λ_j . Then the i.i.d. representation of $\widehat{P}_j(s, t) - P_j(s, t)$ comes from that of F_n^{KM} (Zhou & Yip 1999) and $\widehat{\Lambda}_j^{\text{NA}}$ (Lo, Mack & Wang 1989).

Proof of Corollary 1. Applying the strong uniform consistency results for the KM and NA estimators (Zhou & Yip 1999, and Lo *et al.* 1989), the proof is straightforward.

Proof of Theorem 2. Recall the definition of the risk set Y_j in (3) and the FRS Y_j^f in (4). Then

$$\frac{1}{n} (Y_j(t) - Y_j^f(t)) = \frac{1}{n} \sum_{i=1}^n (\mathbf{1}(X_i^* = j) - P_j(T_i, \infty)) \mathbf{1}(T_i \geq t, \delta_i = 0) + \frac{1}{n} \sum_{i=1}^n (P_j(T_i, \infty) - \widehat{P}_j(T_i, \infty)) \mathbf{1}(T_i \geq t, \delta_i = 0). \quad (14)$$

The first term is a normalized sum of zero-mean summands (see Datta & Satten 2000, lemma 3.1). The second term in (14) can be written as follows:

$$\begin{aligned} & - \int_t^{b_H} (\widehat{P}_j(v, \infty) - P_j(v, \infty)) d\overline{H}_n^c(v) \\ & = - \frac{1}{n} \sum_{i=1}^n \int_t^{b_H} \zeta_j(T_i, X_i, \delta_i, v, \infty) d\overline{H}^c(v) + s_n(t), \end{aligned}$$

where $\overline{H}_n^c$ is the empirical estimator of $\overline{H}^c$ in (9), ζ_j is given in (13), and $\sup_{0 \leq t \leq b_H} |s_n(t)| = O(n^{-1} \ln n)$.

Proof of Corollary 2. The triangle inequality allows us to bound the first term in the decomposition (14) by the sum

$$\left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \geq t, \delta_i = 0) \mathbf{1}(X_i^* = j) - \Pr(T \geq t, \delta = 0, X^* = j) \right| + \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \geq t, \delta_i = 0) - \Pr(T \geq t, \delta = 0) \right| \Pr(X^* = j | \delta = 0).$$

These summands are the absolute error of estimation of some empirical distribution functions. An immediate consequence of the Dvoretzky–Kiefer–Woldfowitz (DKW) bound for empirical measures is that the supremum of the first term in (14) is $O(n^{-1/2} (\ln n)^{1/2})$ almost surely. The second term in (14) can be written as follows:

$$-\int_t^{b_H} (\widehat{P}_j(v, \infty) - P_j(v, \infty)) \mathbf{1}(v \geq t) dH_n^c(v),$$

with H_n^c the empirical estimator of H^c in (9). Then, applying Corollary 1, the supremum of the second term in (14) is $o(n^{-1/2} \ln n)$ almost surely.

The outline of the proof of Theorem 3 is similar to that of Major & Rejtö (1988, theorem 1). We start with a few preliminary results that will be useful for some methods in the main body of the proof of Theorem 3 and Corollary 3. Consider the function $\bar{H}_j$ in (11) and its empirical estimator

$$\bar{H}_{nj}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \geq t, X_i^* = j).$$

The following lemmas give some consistency results for $\bar{H}_{nj}$.

Lemma 1. $\sup_{0 \leq t \leq b_H} |\bar{H}_j(t) - \bar{H}_{nj}(t)| = O(n^{-1/2} (\ln n)^{1/2})$ a.s.

Proof. It is an immediate consequence of the DKW bound for empirical measures.

Lemma 2. $\sup_{k: T_k \leq b_H} \bar{H}_j(T_k) / \bar{H}_{nj}(T_k) = O(\ln n)$ a.s.

Proof. The proof follows the same steps, in the two-dimensional case, as in Shorack & Wellner (1986, p. 415).

Proof of Theorem 3. From the definition of Λ_j^* in (2) and the representation (7) of the FRS estimator, we have

$$\widehat{\Lambda}_j^{*f}(t) - \Lambda_j^*(t) = \int_0^t \frac{d(H_{nj}^{nc}(v) - H_j^{nc}(v))}{\bar{H}_j(v)} - \int_0^t \frac{\bar{H}_{nj}(v) - \bar{H}_j(v)}{\bar{H}_j^2(v)} dH_j^{nc}(v) + R_{j1}(t) + R_{j2}(t) + R_{j3}(t), \quad (15)$$

where H_{nj}^{nc} and $\bar{H}_{nj}$ are the empirical estimators of H_j^{nc} and $\bar{H}_j$ given in (10), and

$$\begin{aligned} R_{j1}(t) &= \int_0^t \frac{\bar{H}_j(v) - \bar{H}_{nj}(v)}{\bar{H}_j^2(v)} d(H_{nj}^{nc}(v) - H_j^{nc}(v)), \\ R_{j2}(t) &= \int_0^t \frac{(\bar{H}_j(v) - \bar{H}_{nj}(v))^2}{\bar{H}_{nj}(v) \bar{H}_j^2(v)} dH_{nj}^{nc}(v), \\ R_{j3}(t) &= \int_0^t \mathbf{1}(Y_j(v) > 0) \left(\frac{1}{Y_j^f(v)} - \frac{1}{Y_j(v)} \right) dN_j(v). \end{aligned}$$

The term R_{j1} can be decomposed into four terms:

$$R_{j1}(t) = \int_0^t \frac{dH_{nj}^{\text{nc}}(v)}{\bar{H}_j(v)} - \int_0^t \frac{\bar{H}_{nj}(v)}{\bar{H}_j^2(v)} dH_{nj}^{\text{nc}}(v) - \int_0^t \frac{dH_j^{\text{nc}}(v)}{\bar{H}_j(v)} + \int_0^t \frac{\bar{H}_{nj}(v)}{\bar{H}_j^2(v)} dH_j^{\text{nc}}(v). \quad (16)$$

The second integral in (16) is a V -statistic of order two. We work with it as follows:

$$\int_0^t \frac{\bar{H}_{nj}(v)}{\bar{H}_j^2(v)} dH_{nj}^{\text{nc}}(v) = \frac{n-1}{n} \int_0^t \frac{dH_{nj}^{\text{nc}}(v)}{\bar{H}_j(v)} + \frac{n-1}{n} \int_0^t \frac{\bar{H}_{nj}(v)}{\bar{H}_j^2(v)} dH_j^{\text{nc}}(v) - \frac{n-1}{n} \int_0^t \frac{dH_j^{\text{nc}}(v)}{\bar{H}_j(v)} + Q_n(t), \quad (17)$$

where we split up the integral into two terms, its diagonal and off-diagonal parts, and obtain the Hájek projection of the U-statistic. The term $Q_n(t)$ is a degenerate U-statistic of order two, and then (see Serfling 1980, section 5.3.3), for each $\delta > 3/2$,

$$\sup_{0 \leq t \leq b_H} |Q_n(t)| = o(n^{-1} \ln n) \text{ almost surely.}$$

Application of the strong law of large numbers (SLLN) to each of the remaining processes in (17) allows us to replace $(n-1)/n$ by 1, so that from (16) we have

$$\sup_{0 \leq t \leq b_H} |R_{j1}(t)| = o(n^{-1} \ln n) \text{ with probability 1.} \quad (18)$$

An immediate consequence of Lemmas 1 and 2 and the SLLN is that

$$\sup_{0 \leq t \leq b_H} |R_{j2}(t)| = O(n^{-1} (\ln n)^2) \text{ with probability 1.} \quad (19)$$

For R_{j3} , consider the decomposition

$$\begin{aligned} R_{j3}(t) = & n^{-1} \int_0^t \frac{Y_j(v) - Y_j^f(v)}{\bar{H}_j^2(v)} dH_j^{\text{nc}}(v) \\ & + n^{-1} \int_0^t (Y_j(v) - Y_j^f(v)) \left(\frac{1}{n^{-2} Y_j^f(v) Y_j(v)} - \frac{1}{\bar{H}_j^2(v)} \right) dH_j^{\text{nc}}(v) \\ & + n \int_0^t \frac{Y_j(v) - Y_j^f(v)}{Y_j^f(v) Y_j(v)} d(H_{nj}^{\text{nc}} - H_j^{\text{nc}})(v). \end{aligned} \quad (20)$$

The last term in (20) is, by Corollary 2, Lemma 2 and the DKW bound for empirical measures, $O(n^{-1} (\ln n)^3)$ almost surely. On applying Corollary 2, Lemma 1 and the DKW bound to the second integral in (20), it is $O(n^{-1} \ln n)$ almost surely. For the first integral in (20), the dominant term comes from the result in Theorem 2,

$$\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\rho_j(T_i, X_i^*, \delta_i, v)}{\bar{H}_j^2(v)} dH_j^{nc}(v) + \int_0^t \frac{s_n(v)}{\bar{H}_j^2(v)} dH_j^{nc}(v),$$

with $\sup_{0 \leq t \leq b_H} |s_n(t)| = O(n^{-1} \ln n)$ almost surely. Therefore,

$$R_{j3}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\rho_j(T_i, X_i^*, \delta_i, v)}{\bar{H}_j^2(v)} dH_j^{nc}(v) + R_{j4}(t)$$

with

$$\sup_{0 \leq t \leq b_H} |R_{j4}(t)| = O(n^{-1} \ln n) \text{ almost surely} \quad (21)$$

The proof is completed using the decomposition (15) and the rates (18), (19) and (21).

Proof of Corollary 3. Consider the decomposition

$$\widehat{\Lambda}_j^{*f}(t) - \Lambda_j^*(t) = \int_0^t \left(\frac{1}{\bar{H}_{nj}(v)} - \frac{1}{\bar{H}_j(v)} \right) dH_{nj}^{nc}(v) + \int_0^t \frac{d(H_{nj}^{nc}(v) - H_j^{nc}(v))}{\bar{H}_j(v)} + \int_0^t \left(\frac{1}{Y_j^f(v)} - \frac{1}{Y_j(v)} \right) dN_j(v). \quad (22)$$

It follows, from Lemmas 1 and 2 and the SLLN, that the absolute value of the first term in (22) is $O(n^{-1/2}(\ln n)^{3/2})$ almost surely.

For the second term in (22), we apply integration by parts and the DKW bound for empirical measures. Then, this term is $O(n^{-1/2}(\ln n)^{1/2})$ almost surely.

The third term in (22) can be written as follows:

$$n \int_0^t \frac{Y_j(v) - Y_j^f(v)}{n^{-2} Y_j^2(v)} dH_{nj}^{nc}(v) - n \int_0^t \frac{Y_j(v) - Y_j^f(v)}{n^{-2} Y_j^2(v) Y_j(v)} \left(\frac{Y_j^f(v)}{Y_j(v)} - 1 \right) dH_{nj}^{nc}(v). \quad (23)$$

The second term in (23) is negligible with respect to the first one. Therefore, from Corollary 2, Lemma 2 and the SLLN, the second term in (23) is $O(n^{-1/2} \ln n)$ almost surely.

References

- Aalen O. Nonparametric inference for a family of counting processes. *Ann. Statist* 1978;6:701–726.
- Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, N. Springer Series in Statistics. New York: Springer-Verlag; 1993. Statistical Models Based on Counting Processes.
- Andrews, DF.; Herzberg, AM. Data: A Collection of Problems From Many Fields for the Student and Research Worker. New York: Springer-Verlag; 1985.
- Bandyopadhyay D, Datta S. Testing equality of survival distributions when the population marks are missing. *J. Statist. Plann. Inference* 2008;138:1722–1732.
- Byar DP, Green SB. The choice of treatment for cancer patients based on covariate information: Applications to prostate cancer. *Bulletin Cancer, Paris* 1980;67:477–488.

- Cheng SC, Fine J, Wei LJ. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* 1998;54:219–228. [PubMed: 9544517]
- Crowder, MJ. *Classical Competing Risks*. London: Chapman and Hall; 2001.
- Datta S, Satten GA. Estimating future stage entry and occupation probabilities in a multistage model based on randomly right-censored data. *Statist. Probab. Lett* 2000;50:89–95.
- Dewan I, Deshpande JV, Kulathinal SB. On testing dependence between time to failure and cause of failure via conditional probabilities. *Scand. J. Statist* 2004;31:79–91.
- Dykstra R, Kochar SC, Robertson T. Restricted tests for testing independence of time to failure and cause of failure in a competing risks model. *Canad. J. Statist* 1998;26:57–68.
- El-Nouty C, Lancar R. The presmoothed Nelson–Aalen estimator in the competing risk model. *Comm. Statist. Theory Methods* 2004;33:135–151.
- Escarela G, Carrière JF. Fitting competing risks with an assumed copula. *Stat. Methods Med. Res* 2003;12:333–349. [PubMed: 12939100]
- Fleming TR. Nonparametric estimation for nonhomogenous Markov processes in the problem of competing risks. *Ann. Statist* 1978;6:1057–1070.
- Kochar SC, Proschan F. Independence of time and cause of failure in the multiple dependent competing risks model. *Statist. Sinica* 1991;1:295–299.
- Lindqvist, BH. *Competing Risks*. In: Ruggeri, F.; Kenett, R.; Faltin, F., editors. *Encyclopedia of Statistics in Quality and Reliability*. New York: Wiley; 2006.
- Lo SH, Mack YP, Wang JL. Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probab. Theory Related Fields* 1989;80:461–473.
- Major P, Rejtö L. Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist* 1988;16:1113–1132.
- Satten GA, Datta S. Kaplan–Meier representation of competing risk estimates. *Statist. Probab. Lett* 1999;42:299–304.
- Serfling, RJ. *Approximation Theorems of Mathematical Statistics*. New York: Wiley; 1980.
- Shorack, G.; Wellner, J. *Empirical Processes with Applications to Statistics*. New York: Wiley; 1986.
- Zhou Y, Yip P. A strong representation of the product-limit estimator for left truncated and right censored data. *J. Mult. Analysis* 1999;69:261–280.

Figure 1.

Fractional risks set (FRS) estimators of Λ_j^* (lines with empty boxes, $j = 1$, and with solid boxes, $j = 2$) and Nelson–Aalen (NA) estimator of Λ (plain line) for Model 1 (under independence of T^* and X^*) and Model 2 (when T^* and X^* are dependent).

Figure 2.

Mean squared error (MSE) of the fractional risk set (FRS) estimators of Λ_j^* (lines with empty boxes, $j = 1$, and with solid boxes, $j = 2$) and Nelson–Aalen (NA) estimator of Λ (plain line, $j = 1$, and dashed line, $j = 2$) when T^* and X^* are independent (Model 1) and dependent (Model 2).

Figure 3. Values of the Kolmogorov–Smirnov test T_{KS} for Model 1 (under H_0 , T^* and X^* are independent) and Model 2 (under H_1).

Figure 4. Fractional risk-set (FRS) estimators of the distribution functions for the two competing risks, prostate cancer ($j = 1$) and other causes ($j = 2$), in the Prostate cancer data set.

Figure 5. Fractional risk-set (FRS) and Nelson–Aalen (NA) estimators of the cumulative hazard functions for the two competing risks, prostate cancer ($j = 1$) and other causes ($j = 2$), in the Prostate cancer data set.