Published in final edited form as:

Lifetime Data Anal. 2014 October ; 20(4): 514–537. doi:10.1007/s10985-013-9280-6.

Nonparametric estimation with recurrent competing risks data

Laura L. Taylor and

Elon University, Campus Box 2320, Elon, NC 27244, USA, ltaylor18@elon.edu

Edsel A. Peña

216 LeConte College, University of South Carolina, Columbia, SC 29208, USA,
pena@stat.sc.edu

Abstract

Nonparametric estimators of component and system life distributions are developed and presented for situations where recurrent competing risks data from series systems are available. The use of recurrences of components' failures leads to improved efficiencies in statistical inference, thereby leading to resource-efficient experimental or study designs or improved inferences about the distributions governing the event times. Finite and asymptotic properties of the estimators are obtained through simulation studies and analytically. The detrimental impact of parametric model misspecification is also vividly demonstrated, lending credence to the virtue of adopting nonparametric or semiparametric models, especially in biomedical settings. The estimators are illustrated by applying them to a data set pertaining to car repairs for vehicles that were under warranty.

Keywords

Recurrent events; Competing risks; Perfect and partial repairs; Martingales; Survival analysis; Repairable systems; Nonparametric methods

1 Introduction

A Q -component competing risks system, also called a series system, is a system which fails when one of its components fails. Consider the situation where such a system can be monitored over the period $[0, \tau]$, where τ is possibly random. In the classical competing risks model, upon system failure the failed component is identified and system monitoring ceases, even if system failure occurred well before time τ . Denoting by T_q , $q = 1, 2, \dots, Q$, the component lifetimes and by $S = \min_q T_q$ the system life, the observable vector from such an experiment or study could be represented by (Z, δ) , where $\delta = 0$ indicates that $S > \tau$ and $Z = \tau$, whereas $\delta = q$ indicates that $S \leq \tau$ and $Z = S = T_q$. With n systems monitored, the observables in this classical competing risks set-up are $\mathbf{D} = \{(Z_i, \delta_i), i = 1, 2, \dots, n\}$. If we assume that these systems are stochastically identical and independent, and if furthermore it is assumed that the T_q 's are independent with T_q having survivor function $F_q(t) = \Pr\{T_q > t\}$

and τ having survivor function $\bar{G}(t) = \Pr\{\tau > t\}$, then inference about $(F_q, q = 1, 2, \dots, Q)$ and $\bar{G}$ are based on $\mathbf{D}$, as well as those for the system life distribution $\bar{F}_S(t) = \prod_{q=1}^Q \bar{F}_q(t)$. This design, however, is not fully efficient since it does not completely utilize available resources. For instance, if each of the n systems fail well before their corresponding τ 's, then the remaining monitoring times $\tau_i - S_i$, $i = 1, 2, \dots, n$, will not be utilized, thus foregoing the potential of gaining more information which may improve inferences about the F_q 's and F_S .

This paper is for the purpose of exploring more resource-efficient options. We will consider the experimental design or study, where upon system failure before the end of the monitoring period at time, τ , either the series system is replaced by an identical series system, referred to as a perfect repair model; or the failed component in the system is replaced by a new component, referred to as a partial repair model (see Lindqvist 2006). In either scenario, system monitoring continues until calendar time τ . In such designs, times to recurrences of component failures will be observed, and the additional information accruing from these recurrences could potentially improve inferences about the distributions, F_q 's. This model has been considered in Taylor and Peña (2013), but only in the situation where the F_q 's are assumed to belong to parametric families of distributions. These methods are related to those based on testing with replacement designs and other variations discussed in Andersen et al. (1993). Additionally, other authors have considered other types of models, such as Cook and Lawless (2007) and Prentice et al. (1981). In the present paper we will consider the nonparametric situation where the F_q 's and G are simply assumed to belong to the space of continuous distribution functions. We remark that parametric assumptions on the F_q 's are more appropriate in engineering and reliability applications, whereas nonparametric or semiparametric models for the F_q 's are more suitable and compelling for biomedical and public health applications. For example, in modeling recurrences of benign and malignant tumors in cancer patients, imposing parametric assumptions on the F_q 's may lead to incorrect conclusions due to misspecified models.

Recurrent competing risks models have not yet been studied extensively, though it has been mentioned in the context of analyzing recurring failures of internal shunts in Cook and Lawless (2007) and Lawless et al. (2001). Many approaches for recurrent events have modeled the time to recurrence, either calendar or gap-time, using conditional or marginal approaches. For example, when there is only one type of event, Prentice et al. (1981) modeled the time of the k th event occurrence via a Cox proportional-hazards regression model. Additionally, Wei et al. (1989) also used the Cox proportional hazards model to marginally model the time to the k th event occurrence. The main difference in their approaches is in the manner in which the at-risk processes are defined. A recent comprehensive review of models and methods for analyzing recurrent events data is given by Geffray (2013).

The aforementioned paper by Lawless et al. (2001) and also Cook and Lawless (2007) present several methods for analyzing the recurrence of failures of internal shunts under a perfect repair. The first analysis applies marginal methods to model the time to failure for the first three recurrences. Additionally, Cook and Lawless (2007) model the time to failure for the first three recurrences due to three forms of failure—obstruction, infection, and

“other.” Each of these analyses modeled failures using a Cox proportional hazards model, and hence the analyses were able to incorporate covariates. Other approaches to dealing with multiple-type events have been explored, such as in Huang and Wang (2005) which dealt with a model with two alternating recurrences; semi-competing risks with a non-terminal and terminal event (see Fine et al. 2001); as well as multi-state models (see Xu et al. 2010).

Ghosh and Lin (2000) presented nonparametric estimation techniques for the mean number of events for recurrent events with the consideration of death as an additional risk. This brings to light the need to consider multiple causes of failures in recurrent event analyses. Dauxois and Sencey (2009) utilize Ghosh and Lin’s results and present nonparametric tests for the recurrent competing risks case regarding the mean number of events. Here we see the identification of multiple types of infections that are observed over time. In this recurrent competing risks scenario, the goal was to compare two types of infections. Chen and Cook (2004) also take death as a terminal event into consideration in their analysis of mean number of events for recurrent events. Each of these papers, and many others, are concerned with the estimation and comparison of the mean number of events between two treatments. The methods in the current paper, on the contrary, are concerned with nonparametrically estimating the marginal distribution functions associated with the inter-event times for two or more recurring competing risks.

Most of the papers cited above have used biological, medical and public health data. In these cases, many of the data will be observational and not from designed experiments. Indeed, even the example considered in this paper is observational. However, the methods developed for recurrent competing risks models in this paper will have many potential applications, such as in reliability and engineering scenarios where component repairs are usually undertaken. Furthermore, in reliability settings, more accurate knowledge of times between failures, or the distributions of these times, of the components constituting the system could be of crucial importance in pinpointing better investment and/or resource allocation decisions to aid in the development of reliable systems, which could only have a positive impact for our society.

The contents of this paper are as follows. The recurrent competing risks model, its notation, and counting processes are presented in Sect. 2. Section 3 presents the derivation of the nonparametric estimators and their associated convergence properties. An application to a small data set concerning the repairs of cars under warranty and comparisons to the parametric methods of Taylor and Peña (2013) are shown in Sect. 5. Small sample properties of the estimator are demonstrated in Sect. 6 as well as an example to demonstrate the advantage nonparametric estimation has over parametric estimation when the underlying model is misspecified. Concluding remarks are given in Sect. 7.

2 Recurrent competing risks model

In this section we describe the statistical model for the recurrent competing risks setting. To achieve generality and conciseness, we shall employ a stochastic process formulation. We first consider a single competing risks system with Q components which will be under observation during the random period $[0, \tau]$. For each $q \in \{1, 2, \dots, Q\}$, we let

$N_q^\dagger = \{N_q^\dagger(s) : s \geq 0\}$ denote the counting process for event occurrences of type q , and let $\mathbf{N}^\dagger = (N_1^\dagger, \dots, N_Q^\dagger)^T$ denote the associated multivariate counting process. Thus, for each q , $N_q^\dagger$ has sample paths that are nondecreasing right-continuous step functions whose jumps are $+1$. For a counting process $N^\dagger$ its associated jump process will be denoted by $N^\dagger$, that is, $N^\dagger(s) = N^\dagger(s) - N^\dagger(s-)$. We shall assume that $\Delta N_q^\dagger \Delta N_{q'}^\dagger = 0$ for every $q \neq q'$, coinciding with the assumption that the interevent times for each of the Q event types are continuous random variables. We let $N_\bullet^\dagger = \sum_{q=1}^Q N_q^\dagger$, the counting process representing event occurrences of all types, and let the at-risk process be $Y^\dagger = \{Y^\dagger(s) : s \geq 0\}$, so that $Y^\dagger(s) = I\{\tau \geq s\}$. In the cases where an additional subscript of i is introduced into the notation, $i = 1, 2, \dots, n$ represents each unit or subject.

We also introduce an effective age process (see Peña et al. 2007) $\mathcal{E}_q = \{\mathcal{E}_q(s) : s \geq 0\}$ and denote $\mathcal{E} = (\mathcal{E}_1, \dots, \mathcal{E}_Q)^T$ the vector of effective age processes. Define $S_0 = 0$ and $S_j = \inf\{s > S_{j-1} : \Delta N_\bullet^\dagger(s) = 1\}$ for $j = 1, 2, \dots, K$, where $K = N_\bullet^\dagger(\tau)$. The S_j 's are the calendar times of occurrences of any type of event. Then, the basic assumption about the effective age processes is that, for each q and each $(S_{j-1}, S_j]$, $\mathcal{E}_q$ is a process whose sample paths are nonnegative, increasing, left-continuous, and differentiable on (S_{j-1}, S_j) . It is also assumed that these effective age processes are completely observable over $[0, \tau]$. Analogously to the S_j 's, for each q , let us also define $S_{q0} = 0$ and $S_{qj} = \inf\{s > S_{qj-1} : \Delta N_q^\dagger(s) = 1\}$ for $j = 1, 2, \dots, K_q$, where $K_q = N_q^\dagger(\tau)$. These S_{qj} 's are the calendar times of occurrences of events of type q . For brevity, we shall also use the following notation:

$$S_{N_\bullet^\dagger(s-)+1} \equiv \min(s, \tau) \text{ and } S_{qN_q^\dagger(s-)+1} \equiv \min(s, \tau), q=1, \dots, Q.$$

We are now in position to present the two effective age processes of interest in this paper. The first is the perfect repair effective age process where

$$\mathcal{E}_q^E(s) = s - S_{N_\bullet^\dagger(s-)}, q=1, 2, \dots, Q; \quad (1)$$

and the second is the partial repair effective age process where

$$\mathcal{E}_q^A(s) = s - S_{qN_q^\dagger(s-)}, q=1, 2, \dots, Q. \quad (2)$$

The perfect repair effective age process coincides with an experiment where the whole system is replaced by an identical new system each time it fails; whereas the partial repair effective age process corresponds to those studies where, when the system fails, only the failed component is replaced by an identical new component. The terminology regarding these two special effective age processes is consistent with those utilized in Bedford and Lindqvist (2004).

For the series system, the observable entity is therefore

$$\mathbf{D} = \left\{ (\mathbf{N}^\dagger(s), \mathcal{E}(s), Y^\dagger(s)) : s \geq 0 \right\}.$$

Technically, $(\mathbf{N}^\dagger(s), \mathcal{E}(s))$ will not be observed for $s > \tau$, but this will not be a problem in the development since their values for $s > \tau$ will not be needed. There is also an associated natural filtration $\mathfrak{F} = \{\mathcal{A}(s) : s \geq 0\}$, where $\mathcal{A}(s) = \sigma\{(\mathbf{N}^\dagger(v), \mathcal{E}(v), Y^\dagger(v)) : v \leq s\}$.

Next we describe the statistical model. We shall assume that τ has distribution function G and that there are distribution functions F_q , $q = 1, 2, \dots, Q$, with associated hazard rate function and cumulative hazard function $\lambda_q(\cdot)$ and $\Lambda_q(\cdot)$, respectively. The reader is referred to Taylor and Peña (2013) for a brief discussion on the relationship between λ_q and the cause-specific hazard function. The distribution functions G and $(F_1, \dots, F_Q)$ will be unknown but will be assumed to belong to the space of absolutely continuous distributions on $[0, \infty)$ with no shared parameters. For each q , define $A_q^\dagger = \{A_q^\dagger(s) : s \geq 0\}$, where

$$A_q^\dagger(s) = \int_0^s Y^\dagger(v) \lambda_q[\mathcal{E}(v)] dv,$$

and let $\mathbf{A}^\dagger = (A_1^\dagger, \dots, A_Q^\dagger)^T$. Next, define $\mathbf{M}^\dagger = (M_1^\dagger, \dots, M_Q^\dagger)^T$, where

$M_q^\dagger = \{M_q^\dagger(s) : s \geq 0\}$ with $M_q^\dagger(s) = N_q^\dagger(s) - A_q^\dagger(s)$. The statistical model governing the competing risks system specifies that $\mathbf{M}^\dagger$ is a Q -vector of zero-mean square-integrable martingales with predictable quadratic variation matrix process given by

$$\langle \mathbf{M}^\dagger, \mathbf{M}^\dagger \rangle(s) = \text{Dg}(\mathbf{A}^\dagger(s)),$$

where $\text{Dg}(\mathbf{a})$ is the diagonal matrix with diagonal elements being the elements of the vector $\mathbf{a}$. For the purpose of making inferences about $(F_1, \dots, F_Q)$ or some functional of interest such as the system life distribution, we shall assume that we have n independent and identically distributed (IID) copies of $\mathbf{D}$ given by $(\mathbf{D}_1, \dots, \mathbf{D}_n)$. The natural filtration of interest will then be $\mathcal{F}(s) = \mathcal{V}_{i=1}^n \mathcal{F}_i(s)$ where $\mathcal{F}_i(s)$ is the filtration generated by data $\mathbf{D}_i$.

In this recurrent competing risks model, T_{ij} will denote the inter-event or gap-time associated with the j th recurrence of the event of type q for the i th unit, where $i = 1, 2, \dots, n$, $q = 1, 2, \dots, Q$, and $j = 1, 2, \dots, K_i$. K_i represents the random number of recurrences observed for system i prior to the monitoring time, τ_i . For brevity, we let $K_{iq} = N_q^\dagger(\tau_i)$, the number of recurrences of the event of type q for unit i . Furthermore, we denote by S_{ij} the calendar time in which the j th recurrence of the event of type q for unit i occurred. In the case of partial repairs, note that $S_{ij} = \sum_{k=1}^j T_{iqk}$. For each cause q , as i and j vary, it is assumed that T_{ij} 's are IID from some unknown distribution F_q .

As an example, consider the data set discussed in Sect. 5. A plot of the data is given in Fig. 1 where each horizontal line represents a car and the symbols represent one of two types of failures. The x axis represents the mileage of the vehicle. We consider a partial repair strategy where the inter-event times for a failure mode or type are given by the lengths of the line segments between two consecutive symbols of the same color.

3 Nonparametric estimators

In this section, we consider the estimation of $(\Lambda_1, \dots, \Lambda_Q)$ and $(F_1, \dots, F_Q)$ based on the data $(\mathbf{D}_1, \dots, \mathbf{D}_n)$. We will also present an estimator of the system life distribution, F_S , of the competing risks system. Our basic approach will follow the now well-trodden route of Andersen et al. (1993), which is to first obtain an estimator of $(\Lambda_1, \dots, \Lambda_Q)$, and to then employ the product-integral relation between Λ_q and F_q to arrive at an estimator of F_q . Let us first focus again on one system. Following Peña et al. (2001), let us define

$Z_q = \{Z_q(s, t) : (s, t) \in \mathcal{R}_+^2\}$ where $Z_q(s, t) = I\{\mathcal{E}_q(s) \leq t\}$. We then define the doubly-indexed process $N_q = \{N_q(s, t) : (s, t) \in \mathcal{R}_+^2\}$, $A_q = \{A_q(s, t) : (s, t) \in \mathcal{R}_+^2\}$, and $M_q = \{M_q(s, t) : (s, t) \in \mathcal{R}_+^2\}$, where for $q = 1, \dots, Q$,

$$N_q(s, t) = \int_0^s Z_q(v, t) N_q^\dagger(dv); \quad A_q(s, t) = \int_0^s Z_q(v, t) A_q^\dagger(dv); \quad M_q(s, t) = \int_0^s Z_q(v, t) M_q^\dagger(dv).$$

Also, introduce the doubly-indexed process $Y_q = \{Y_q(s, t) : (s, t) \in \mathcal{R}_+^2\}$, referred to as the generalized at-risk process, according to

$$Y_q(s, t) = \sum_{j=1}^{N_q^\dagger(s-) + 1} \frac{I\{\mathcal{E}_q(S_{j-1}) < t \leq \mathcal{E}_q(S_j)\}}{\mathcal{E}_q'[\mathcal{E}_{qj}^{-1}(t)]}$$

where $\mathcal{E}_{qj}(\cdot)$ is the restriction of $\mathcal{E}_q$ on $(S_{j-1}, S_j]$ with $\mathcal{E}_{qj}^{-1}$ the associated inverse function, and $\mathcal{E}_q'$ is the derivative of $\mathcal{E}_q$. We first present the specialized forms of this generalized at-risk process for the perfect repair and partial repair designs.

Proposition 1 *If the effective age processes are the $\mathcal{E}_q^E$'s given in Eq. (1), then*

$$Y_q(s, t) = \sum_{j=1}^{N_q^\dagger(s-) + 1} I\{T_j \geq t\},$$

where $T_j = S_j - S_{j-1}$ for $j = 1, 2, \dots, K$, and with $T_{K+1} = \tau - S_K$.

Proof This is immediate from the form of the perfect repair effective age process.

Proposition 2 *If the effective age processes are the $\mathcal{E}_q^A$'s given in Eq. (2), then*

$$Y_q(s, t) = \sum_{j=1}^{N_q^\dagger(s-) + 1} I \{T_{qj} \geq t\},$$

where $T_{qj} = S_{qj} - S_{qj-1}$ for $j = 1, 2, \dots, K_q$, and with $T_{qK_q+1} = \tau - S_{qK_q}$.

Proof Fix q . Set $L_0 = 0$ and, for $l=1, \dots, N_q^\dagger(s-) + 1$, let L_l be such that $S_{ql} = S_{L_l}$. Then we have

$$Y_q(s, t) = \sum_{l=1}^{N_q^\dagger(s-) + 1} \sum_{j=1}^{L_l - L_{l-1}} I \left\{ \mathcal{E}_q(S_{L_{l-1}+j-1}) < t \leq \mathcal{E}_q(S_{L_{l-1}+j}) \right\}.$$

But now from the definition of the partial repair effective age process, we find that, for each $l=1, \dots, N_q^\dagger(s-) + 1$, we have

$$\sum_{j=1}^{L_l - L_{l-1}} I \left\{ \mathcal{E}_q(S_{L_{l-1}+j-1}) < t \leq \mathcal{E}_q(S_{L_{l-1}+j}) \right\} = I \{T_{ql} \geq t\},$$

completing the proof of the proposition.

The importance of this generalized at-risk process is the identity given by

$$A_q(s, t) = \int_0^t Y_q(s, w) \Lambda_q(dw), \quad q=1, \dots, Q, \quad (3)$$

which is easily established by following the proof of an analogous result in Peña et al. (2007), see also Peña et al. (2001). Note that the argument of $\lambda_q(\cdot)$ in the representation in Eq. (3) is just w , whereas in its original form it is the random process $\mathcal{E}_q(v)$. But then noting that for fixed t , the process $M_q(\cdot, t)$ is still a martingale, it follows that

$$E \{N_q(s, t)\} = E \left\{ \int_0^t Y_q(s, w) \Lambda_q(dw) \right\}. \quad (4)$$

With the n IID copies $\mathbf{D}_i$, $i = 1, \dots, n$, of $\mathbf{D}$, define now the processes N_{iq} and Y_{iq} for $i = 1, \dots, n$; $q = 1, \dots, Q$, and the aggregated processes

$$N_{\bullet q} = \sum_{i=1}^n N_{iq} \text{ and } Y_{\bullet q} = \sum_{i=1}^n Y_{iq}.$$

Introduce the indicator process $J_{\bullet q} = \{J_{\bullet q}(s, t) : (s, t) \in \mathcal{R}_+^2\}$ with $J_{\bullet q}(s, t) = I \{Y_{\bullet q}(s, t) > 0\}$. Then, from the moment identity in Eq. (4), a Nelson–Aalen type estimator of $\Lambda_q(t)$ is provided by

$$\hat{\Lambda}_q(s, t) = \int_0^t \frac{J_{\bullet q}(s, w)}{Y_{\bullet q}(s, w)} N_{\bullet q}(s, dw). \quad (5)$$

Through the product-integral representation of $\bar{F}_q = 1 - F_q$ in terms of Λ_q , we obtain its product-limit type estimator via

$$\hat{\bar{F}}_q(s, t) = \prod_{w=0}^t [1 - \hat{\Lambda}_q(s, dw)].$$

Furthermore, since for a competing risks system the cumulative hazard function of the system life is $\Lambda_s(t) = \sum_{q=1}^Q \Lambda_q(t)$, then an obvious estimator is

$$\hat{\Lambda}_s(s, t) = \sum_{q=1}^Q \hat{\Lambda}_q(s, t).$$

Again, by the product-integral representation, an estimator of the system life survivor distribution $\bar{F}_S$ is given by

$$\hat{\bar{F}}_S(s, t) = \prod_{w=0}^t [1 - \hat{\Lambda}_s(s, dw)]. \quad (6)$$

In the perfect repair case, only when $Q = 1$, $\hat{\bar{F}}_S$ reduces to the estimator of Peña et al. (2001). In the partial repair case, when $Q = 1$, $\hat{\bar{F}}_S$ leads to the estimator in Peña et al. (2007) where there are no covariates. Another reasonable estimator of $\bar{F}_S(t)$ is provided by

$$\tilde{\bar{F}}_S(s, t) = \prod_{q=1}^Q \hat{\bar{F}}_q(s, t). \quad (7)$$

We surmise at this point that the properties of this estimator will be very similar to that of $\hat{\bar{F}}_S(s, t)$ for small to moderate sample sizes and that they will be asymptotically (as $n \rightarrow \infty$) equivalent.

We remark that in the estimators of the Λ_q 's, $\bar{F}_q$'s, Λ_S , and $\bar{F}_S$ presented above, the value of s will usually be set to $s = \max_{i=1, \dots, n} \tau_i$. Also, before we proceed to examine the properties of these estimators, we point out an advantage of the system life distribution estimator that is formed from the component life distribution estimators. First, imagine that the design is a perfect repair design. Then, an estimator of the system life distribution could be obtained via the nonparametric estimator in Peña et al. (2001) using the

$N_{\bullet\bullet} = \sum_{i=1}^n \sum_{q=1}^Q N_{iq}$ and $Y_{\bullet\bullet} = \sum_{i=1}^n \sum_{q=1}^Q Y_{iq}$ processes. On the other hand, consider the

partial repair design. In this case, the system life estimators in Eqs. (6) and (7) are still viable, but now note that we will be unable to obtain a valid estimator of the system life distribution based only on the aggregated processes $N_{\bullet\bullet}$ and $Y_{\bullet\bullet}$. This shows the wider potential of first obtaining estimators of component lifetime distributions and then forming the system lifetime distribution estimator from these estimators.

4 Properties of estimators and some efficiency issues

In this section we present the main asymptotic distributional properties of the estimators and examine some issues of efficiency. Let us denote by $\hat{\Lambda}(s, t) = (\hat{\Lambda}_1(s, t), \dots, \hat{\Lambda}(s, t))^T$ and by $\Lambda(t) = (\Lambda_1(t), \dots, \Lambda_Q(t))^T$. With

$$\mathbf{H}(s, t) = \text{Dg}\left(\frac{J_{\bullet q}(s, t)}{Y_{\bullet q}(s, t)}, q=1, 2, \dots, Q\right),$$

we are able to write in matrix form the vector of estimators via

$$\hat{\Lambda}(s, t) = \int_0^t \mathbf{H}(s, w) \mathbf{N}_{\bullet}(s, dw) \text{ with the obvious notation that}$$

$$\mathbf{N}_{\bullet} = \sum_{i=1}^n \left(N_{i1}, \dots, N_{iQ} \right)^T.$$

Let us also define

$$\Lambda^*(s, t) = \int_0^t \text{Dg}(J_{\bullet q}, q=1, \dots, Q) \Lambda(dw).$$

Then, we have the integral representation,

$$\sqrt{n}[\hat{\Lambda}(s, t) - \Lambda^*(s, t)] = \frac{1}{\sqrt{n}} \int_0^t [n\mathbf{H}(s, w)] \mathbf{M}_{\bullet}(s, dw),$$

where $\mathbf{M}_{\bullet} = \sum_{i=1}^n \left(M_{i1}, \dots, M_{iQ} \right)^T$. Let $(s^*, t^*) \in \mathcal{R}_+^2$ be such that there exists deterministic functions $\{y_q(\cdot, \cdot), q=1, \dots, Q\}$ satisfying, for $s \in (0, s^*]$ and as $n \rightarrow \infty$,

$$\max_{q=1, \dots, Q} \sup_{t \in [0, t^*]} \left| \frac{1}{n} Y_{\bullet q}(s, t) - y_q(s, t) \right| \xrightarrow{p} 0; \quad \inf_{t \in [0, t^*]} y_q(s, t) > 0.$$

Under these conditions, it follows from the weak convergence theorem presented in Peña et al. (2000) that, for $s \in (0, s^*]$ and as $n \rightarrow \infty$,

$$\left\{ \sqrt{n}[\hat{\Lambda}(s, t) - \Lambda^*(s, t)] : t \in [0, t^*] \right\} \Rightarrow \{Z(s, t) : t \in [0, t^*]\},$$

where “ $\Rightarrow$ ” means weak convergence in Skorohod’s $\mathcal{D}[0, t^*]^Q$ -space and $\{\mathbf{Z}(s, t) : t \in [0, t^*]\}$ is a zero-mean Gaussian process with covariance matrix function $\Sigma(s, \cdot)$ given by

$$\Sigma(s, t) = \text{Dg} \left(\int_0^t \frac{\Lambda_q(dw)}{y_q(s, w)}, q=1, \dots, Q \right).$$

Coupling this weak convergence result with the fact that

$$\max_{q=1, \dots, Q} \sup_{t \in [0, t^*]} \left| \Lambda_q^*(s, t) - \Lambda_q(t) \right| \xrightarrow{p} 0,$$

and via the triangle inequality, we have therefore established the following weak convergence result.

Theorem 1 *Under the conditions stated above, as $n \rightarrow \infty$ and for $s \in (0, s^*]$, we have*

$$\left\{ \sqrt{n}[\hat{\Lambda}(s, t) - \Lambda(t)] : t \in [0, t^*] \right\} \Rightarrow \{\mathbf{Z}(s, t) : t \in [0, t^*]\}.$$

From this main weak convergence result, a cascade of other weak convergence results follow via the Functional Delta Method. For instance, we have that

$$\left\{ \sqrt{n}[\hat{\mathbf{F}}(s, t) - \overline{\mathbf{F}}(t)] : t \in [0, t^*] \right\} \Rightarrow \{\text{Dg}(\overline{\mathbf{F}}(t))\mathbf{Z}(s, t) : t \in [0, t^*]\}, \quad (8)$$

so that the asymptotic distribution of $\hat{F}_q(s, t)$ is given by

$$\hat{F}_q(s, t) \sim \text{AN} \left[\overline{F}_q(t), \frac{1}{n} \overline{F}_q(t)^2 \int_0^t \frac{\Lambda_q(dw)}{y_q(s, w)} \right].$$

It also follows from Theorem 1 that

$$\left\{ \sqrt{n}[\hat{\Lambda}^*(s, t) - \Lambda_s(t)] : t \in [0, t^*] \right\} \Rightarrow \left\{ \mathbf{1}^T \mathbf{Z}(s, t) : t \in [0, t^*] \right\}, \quad (9)$$

which consequently leads to the result that

$$\left\{ \sqrt{n}[\hat{F}_s(s, t) - \overline{F}_s(t)] : t \in [0, t^*] \right\} \Rightarrow \left\{ \overline{F}_s(t) \mathbf{1}^T \mathbf{Z}(s, t) : t \in [0, t^*] \right\}.$$

As a consequence we obtain the asymptotic distribution of $\hat{F}_s$ given by

$$\hat{F}_s(s, t) \sim \text{AN} \left[\overline{F}_s(t), \frac{1}{n} \overline{F}_s(t)^2 \sum_{q=1}^Q \int_0^t \frac{\Lambda_q(dw)}{y_q(s, w)} \right].$$

Consistent estimators of the asymptotic variance for $\hat{\bar{F}}_q(t)$ and $\hat{\bar{F}}_s(t)$ are, respectively, given by

$$\hat{\bar{F}}_q(t)^2 \int_0^t \frac{\hat{\Lambda}_q(dw)}{Y_{\bullet q}(s, w)}$$

and

$$\hat{\bar{F}}_s(t)^2 \sum_{q=1}^Q \int_0^t \frac{\hat{\Lambda}_q(dw)}{Y_{\bullet q}(s, w)}. \quad (10)$$

In addition, as surmised earlier, from Eq. (8) and the functional delta method, we do find also that the asymptotic distribution of the alternative estimator $\tilde{\bar{F}}_s(s, t)$ given in Eq. (7) coincides exactly with that of $\hat{\bar{F}}_s(s, t)$, certainly a not-so-surprising result!

Let us now specialize these results to the perfect repair and partial repair designs. We recall a useful result in Peña et al. (2001) which will facilitate our obtaining the $y_q(s, t)$ -functions for the two designs of interest. Let $\mathbf{T} = (T_1, T_2, \dots)$ be a sequence of IID random variables from a continuous distribution F on $[0, \infty)$ and let τ have a continuous distribution G , also defined on $[0, \infty)$. Assume that $\mathbf{T}$ and τ are independent. Let $S_0 = 0$ and $S_k = \sum_{j=1}^k T_j$, and define

$$K(s) = K(s; \mathbf{T}, \tau) = \max \{k \in \{0, 1, 2, \dots\} : S_k \leq \min(s, \tau)\}.$$

Define $Y(s, t) = \sum_{j=1}^{K(s-)} I(T_j \geq t + I \{ \min(s, \tau) - S_{K(s-)} \geq t \})$. Then, Proposition 2 in Peña et al. (2001) states that, with $\vartheta(\cdot) = \sum_{k=1}^{\infty} F^{\star k}(\cdot)$ and $F^{\star k}$ being the k th convolution of F ,

$$E \{Y(s, t)\} \equiv y(s, t) = \bar{F}(t) \bar{G}(t) \left[1 + \frac{1}{\bar{G}(t)} \int_t^s \vartheta(w - t) G(dw) \right] I \{t \leq s\}.$$

Furthermore, if $Y_i(\cdot, \cdot)$, $i = 1, 2, \dots$ are IID copies of $Y(\cdot, \cdot)$, then $\bar{Y}(s, t) = \frac{1}{n} \sum_{i=1}^n Y_i(s, t)$ will converge uniformly in probability to $y(s, t)$ in t for each $s \in \mathcal{R}_+$ as $n \rightarrow \infty$.

Proposition 3 *The uniform-in-probability limit functions of*

$$\bar{Y}_q(s, t) = \frac{1}{n} \sum_{i=1}^n Y_{iq}(s, t)$$

under the perfect repair and partial repair effective age processes are given, respectively, by

$$y_q^E(s, t) = \bar{F}_s(t) \bar{G}(t) \left[1 + \frac{1}{\bar{G}(t)} \int_t^s \vartheta_s(w-t) G(dw) \right] I\{t \leq s\}; \quad y_q^A(s, t) = \bar{F}_q(t) \bar{G}(t) \left[1 + \frac{1}{\bar{G}(t)} \int_t^s \vartheta_q(w-t) G(dw) \right] I\{t \leq s\},$$

where $\vartheta_q(\cdot)$ is the renewal function of F_q and $\vartheta_S(\cdot)$ is the renewal function of F_S .

Proof The proofs of these results follow immediately from the aforementioned result in Peña et al. (2001) and the representations provided in Propositions 1 and 2.

In the special case where $F_q(t) = \exp(-\theta_q t) I\{t \geq 0\}$ for $q = 1, \dots, Q$, and $\bar{G}(t) = \exp(-\eta t) I\{t \geq 0\}$, we are able to get closed-form expressions for the y_q^E and y_q^A functions. With

$\theta_\bullet = \sum_{q=1}^Q \theta_q$, straightforward derivations reveal that

$$y_q^E(s, t) = e^{-(\theta_\bullet + \eta)t} \left\{ \left(1 + \frac{\theta_\bullet}{\eta} \right) - \frac{\theta_\bullet}{\eta} e^{-\eta(s-t)} [1 + \eta(s-t)] \right\} I\{t \leq s\};$$

$$y_q^A(s, t) = e^{-(\theta_q + \eta)t} \left\{ \left(1 + \frac{\theta_q}{\eta} \right) - \frac{\theta_q}{\eta} e^{-\eta(s-t)} [1 + \eta(s-t)] \right\} I\{t \leq s\}.$$

In the limiting case where $s \rightarrow \infty$, these expressions simplify to

$$y_q^E(\infty, t) = e^{-(\theta_\bullet + \eta)t} \left(1 + \frac{\theta_\bullet}{\eta} \right) \text{ and } y_q^A(\infty, t) = e^{-(\theta_q + \eta)t} \left(1 + \frac{\theta_q}{\eta} \right).$$

In this limiting case and with exponential distributions let us compare the efficiencies of the estimator of F_S under the two effective age processes. Simple calculations show that

$$\sigma_E^2(t) \equiv \text{AsyVar} \left\{ \sqrt{n} \hat{F}_S^E(\infty, t) \right\} = e^{-2\theta_\bullet t} \frac{\theta_\bullet \eta}{(\theta_\bullet + \eta)^2} [e^{(\theta_\bullet + \eta)t} - 1];$$

$$\sigma_A^2(t) \equiv \text{AsyVar} \left\{ \sqrt{n} \hat{F}_S^A(\infty, t) \right\} = e^{-2\theta_\bullet t} \sum_{q=1}^Q \frac{\theta_q \eta}{(\theta_q + \eta)^2} [e^{(\theta_q + \eta)t} - 1].$$

We may then ask which of the two designs will lead to a more efficient estimator of the system life distribution, or for what values of t one estimator will be better than the other. Some in-depth thinking about the two designs will lead us to expect that the perfect repair design should lead to an estimator that is better for small values of t , whereas the estimator based on the partial repair should do somewhat better for larger values of t . We should be able to verify these expectations, at least for this special case of exponential distributions. From the expressions above, the asymptotic relative efficiency of the estimator based on the perfect repair relative to the estimator based on the partial repair is

$$\text{ARE}(t) \equiv \frac{\sigma_A^2(t)}{\sigma_E^2(t)} = \frac{\sum_{q=1}^Q \frac{\theta_q}{(\theta_q + \eta)^2} [e^{(\theta_q + \eta)t} - 1]}{\frac{\theta_\bullet}{(\theta_\bullet + \eta)^2} [e^{(\theta_\bullet + \eta)t} - 1]}.$$

Observe that

$$\text{ARE}(0) \equiv \lim_{t \downarrow 0} \text{ARE}(t) = \left[\frac{\theta_{\bullet} + \eta}{\theta_{\bullet}} \right] \sum_{q=1}^Q \left[\frac{\theta_q}{\theta_q + \eta} \right]; \quad (11)$$

whereas, it is easy to see that $\text{ARE}(\infty) \equiv \lim_{t \uparrow \infty} \text{ARE}(t) = 0$. Is $\text{ARE}(0)$ always exceeding unity and does it have an upper bound? The answers to these questions are contained in the following proposition, which confirms our expectations regarding the values of t where the perfect repair and partial repair estimators of the system life distribution will tend to perform well, at least for the exponential F_q 's and G .

Proposition 4 *The expression for $\text{ARE}(0)$ given in Eq. (11) satisfies the inequalities $1 < \text{ARE}(0) < Q$ for all positive θ_q 's and positive η .*

Proof To establish the first inequality, we have that

$$\sum_q \frac{\theta_q}{\theta_q + \eta} \geq \sum_q \frac{\theta_q}{\max_{q'} \theta_{q'} + \eta} = \frac{\theta_{\bullet}}{\max_{q'} \theta_{q'} + \eta} > \frac{\theta_{\bullet}}{\theta_{\bullet} + \eta}$$

since $\theta_{\bullet} > \max_{q'} \theta_{q'}$. To prove the second inequality, consider the mapping $o \mapsto g(o) = o/(o + \eta)$. This is a concave map on $\mathcal{R}_+$. Think now of a random variable O taking the values θ_q , $q = 1, \dots, Q$, each with probability $1/Q$. Applying Jensen's Inequality, we find that

$$E[g(O)] = \frac{1}{Q} \sum_{q=1}^Q \frac{\theta_q}{\theta_q + \eta} < g[E(O)] = g[\theta_{\bullet}/Q] = \frac{\theta_{\bullet}}{\theta_{\bullet} + Q\eta} \leq \frac{\theta_{\bullet}}{\theta_{\bullet} + \eta},$$

completing the proof of the proposition.

We may also inquire regarding the gain in efficiency of estimators that take into account recurrences over estimators that only use the first event times. Let us denote by $\tilde{\bar{F}}_q$ the estimator of F_q which uses only the first event times of the n systems, and by $\tilde{\bar{F}}_s$ the corresponding system life distribution estimator. Then it is easy to see that

$$\tilde{\bar{F}}_s(t) \sim \text{AN}[\bar{F}_s(t), \frac{1}{n} \bar{F}_s(t)^2 \int_0^t \frac{\Lambda_s(dw)}{\bar{F}_s(w) \bar{G}(w)}].$$

If the perfect repair effective age process is employed, the asymptotic relative efficiency of this estimator relative to $\hat{\bar{F}}_s(t)$ is

$$\text{ARE}[\tilde{\bar{F}}_s(t): \hat{\bar{F}}_s^E(s, t)] = \left[\int_0^t \frac{\Lambda_s(dw)}{\bar{F}_s(w) \bar{G}(w) [1 + \nu^*(s, w)]} \right] \left[\int_0^t \frac{\Lambda_s(dw)}{\bar{F}_s(w) \bar{G}(w)} \right]^{-1}$$

where

$$\nu^*(s, w) = \frac{1}{\bar{G}(t)} \int_t^s \vartheta^*(w - t) G(dw) I\{t \leq s\}.$$

For example, if we let $s \rightarrow \infty$ and assume $F_q = \text{EXP}(\theta_q)$ and $G = \text{EXP}(\eta)$, we find that $\nu^*(\infty, w) = \theta_\bullet/\eta$ so that $\text{ARE}[\tilde{F}_s(t): \hat{F}_s^E(\infty, t)] = \eta/(\theta_\bullet + \eta)$.

If instead a partial repair effective age process is utilized, then we get

$$\text{ARE}[\tilde{F}_s(t): \hat{F}_s^A(s, t)] = \left[\sum_{q=1}^Q \int_0^t \frac{\Lambda_q(dw)}{\bar{F}_q(w) \bar{G}(w) [1 + \nu_q(s, w)]} \right] \left[\int_0^t \frac{\Lambda_s(dw)}{\bar{F}_s(w) \bar{G}(w)} \right]^{-1}$$

with

$$\nu_q(s, w) = \frac{1}{\bar{G}(t)} \int_t^s \vartheta_q(w - t) G(dw) I\{t \leq s\}.$$

As in the concrete example above pertaining to exponential F_q 's and G , straightforward calculations yield

$$\text{ARE}[\tilde{F}_s(t): \hat{F}_s^A(\infty, t)] = \left[\frac{(\eta + \theta_\bullet)\eta}{\theta_\bullet} \right] \frac{\sum_{q=1}^Q \left[\frac{\theta_q}{(\eta + \theta_q)^2} \right] [\exp\{(\theta_q + \eta)t\} - 1]}{\exp\{(\theta_\bullet + \eta)t\} - 1}.$$

The right-hand side goes to zero as $t \rightarrow \infty$, while its limit as $t \downarrow 0$ is bounded above by $\eta/(\eta + \min_q \theta_q)$, which is less than unity. That is, $\tilde{F}_s(t)$ is always asymptotically less efficient than $\hat{F}_s^A(\infty, t)$ in this exponential setting.

Motivated by a question by Professor Joseph Gastwirth, we also examine the loss in efficiency by utilizing the nonparametric estimator of the system life distribution relative to a parametric estimator of the system life, provided the parametric model assumptions are valid. From Taylor and Peña (2013) it was found that for the fully parametric estimator of the system life distribution, which we shall denote by $\hat{F}_{S,PARA}$, we have

$$\text{AsyVar} \left\{ \sqrt{n} \hat{F}_{S,PARA}(s, t) \right\} = \bar{F}_s(t)^2 \sum_{q=1}^Q \dot{\Lambda}_q(t; \theta_q) \mathcal{J}_{qq}(s; \theta_q, G)^{-1} \dot{\Lambda}_q(t; \theta_q),$$

where $\dot{\Lambda}_q(t; \theta_q) = \partial \Lambda_q(t; \theta_q) / \partial \theta_q$ and

$$\mathcal{J}_{qq}(s; \theta_q, G) = \int_0^s y_q(s, w; \theta, G) \left[\frac{\partial}{\partial \theta_q} \log \lambda_q(w; \theta_q) \right]^{\otimes 2} \Lambda_q(dw; \theta_q).$$

Consequently, the asymptotic relative efficiency of the fully nonparametric estimator relative to the parametric estimator at time t is

$$\text{ARE} \left\{ \hat{F}_S(s, t) : \hat{F}_{S,PARA}(s, t) \right\} = \left[\sum_{q=1}^Q \int_0^s \frac{\Lambda_q(dw; \theta_q)}{y_q(s, w; \theta_q, G)} \right]^{-1} \left[\sum_{q=1}^Q \dot{\Lambda}_q(t; \theta_q) \mathcal{I}_{qq}(s; \theta_q, G)^{-1} \dot{\Lambda}_q(t; \theta_q) \right].$$

To provide a concrete illustration of this asymptotic relative efficiency, let us assume a perfect repair effective age and let $F_q = \text{EXP}(\theta_q)$, $G = \text{EXP}(\eta)$, and $s \rightarrow \infty$. Direct calculations yield that in this case

$$\text{ARE} \left\{ \hat{F}_S(\infty, t) : \hat{F}_{S,PARA}(\infty, t) \right\} = \frac{(\theta_{\bullet} + \eta)^2 t^2}{\exp \{(\theta_{\bullet} + \eta)t\} - 1}.$$

Observe that as $z \equiv (\theta_{\bullet} + \eta)t$ goes to either zero or infinity, the ARE-value goes to zero, and the maximum ARE is achieved at the z -value satisfying $\exp(z) = 2/(2-z)$, which is about $z_0 = 1.5936$ and with an associated maximum ARE-value of 0.6476. Thus, the fully nonparametric estimator may indeed lose efficiency relative to the parametric estimator, though it should at the same time be pointed out that if the parametric model assumptions do not hold, then the parametric estimator could become very highly inefficient.

5 Application

A data set containing the mileage at time of servicing for 189 cars under warranty were obtained from Professor Ananda Sen of the University of Michigan. The data also indicates both the labor code (indicated by 1–14) and a failure mode (indicated by 1 or 2) at each of these service mileages for the vehicles. For simplicity, we have modeled the data using only the mileages at servicing and the failure modes as the competing recurrent events. This service data has several limitations that should be noted when applying our methods for illustrative purposes. First, the data is discrete, and our methods are designed for continuous data. The data set also does not have any censored values as each period of observation ended with a service. Additionally, in the raw data, some vehicles underwent multiple services at one time so multiple labor codes and failure modes were recorded for a single mileage. When this occurred, only the first listed failure mode was used. Due to the wear and tear of mechanical components, deterioration of the system and its components is most plausible. Therefore, the independence and/or identical distribution of the component lifetimes within a system is most likely violated. A final consideration about this data is that it will not be representative of the population of cars as only cars with warranty claims were reported in this data. Nevertheless, with these caveats, we still use this data set to compare our methods with the parametric analysis performed in Taylor and Peña (2013).

This data was previously used by Taylor and Peña (2013) to demonstrate their parametric method for recurrent competing risks to compare failure modes 1 and 2. In their application, an underlying Weibull distribution was assumed for the inter-event time distribution associated with each of the failure modes operating under a partial repair. For comparisons,

Taylor and Peña (2013) also calculated a product limit estimator using only the first event. The product limit estimator utilizing only the first-event for each car indicated a large discrepancy between this estimator and the parametric method proposed in Taylor and Peña (2013) leading one to question the validity of the assumed Weibull distribution.

Figure 2 updates the graph shown in Taylor and Peña (2013) so that the nonparametric estimate presented in this paper which accounts for recurrent competing risks under a partial repair strategy is displayed. Again, there is much discrepancy between each of these estimates. Several assumptions were made in the application of these methods to this data. For example, it was assumed that the cars were repaired via a partial repair assumption so that only the failed component was replaced with a new component. It could be that the nature of these repairs coincided with other types of maintenance that could have improved the still-functioning components of the cars. However, our analysis assumes that only the non-functioning component was repaired. The mileage at servicing could also vary greatly between the time the issue appeared and the time the car was brought in for repairs. However it is worth noting that the parametric model assumes underlying Weibull inter-event time distributions, and the product limit estimator is limited by only observing the first failure. While the Weibull distribution is fairly flexible in the shapes it can form, there is no verification that this model is appropriate for this data. Moreover, the nonparametric model presented in this paper makes no distributional assumptions (as opposed to the parametric model from Taylor and Peña 2013) and utilizes all observations in the data set (as opposed to the product limit estimator shown). Point-wise 95 % confidence bands were also generated based on Eq. (10) and are depicted in Fig. 2 as well.

6 Simulation studies

6.1 Demonstration with simulated data

Inter-event times were randomly generated from Weibull distributions for two competing risks with parameters $(\alpha_1 = 2, \beta_1 = 1/2)$ and $(\alpha_2 = 3, \beta_2 = 1/3)$. The randomly generated data is displayed in Fig. 3 and contains data from 10 units. The monitoring times followed an exponential distribution with mean one. Summary statistics for the randomly generated data are given in Table 1. The nonparametric recurrent competing risk estimate for the cumulative hazard is shown in Fig. 4 for the data generated in Fig. 3. The red line represents the theoretical cumulative hazard function and the dotted blue line represents the parametric recurrent competing risks estimate from Taylor and Peña (2013) correctly assuming underlying Weibull distributions. To illustrate the issues with model misspecification when employing parametric models, it was assumed in Fig. 5 that the underlying inter-event time distributions were from an exponential distribution. In this case, the estimate of the system lifetime distribution under the methods developed under parametric models in Taylor and Peña (2013) leads to a grossly incorrect estimate of the system lifetime distribution when compared to the nonparametric estimate of the system lifetime distribution based on the system lifetime distribution estimator discussed in this paper.

6.2 Simulation to examine properties

Data generation of 10 units with nonparametric estimation of the inter-event time distribution were repeated for 1,000 iterations. Figure 6 displays the mean nonparametric estimate calculated at every fifth percentile from the fifth through 95th percentiles for the nonparametric estimator of the cumulative hazard (black, solid) and the theoretical cumulative hazard curve based on the Weibull inter-event times (red, dashed). Additionally, the simulated bias and mean-squared error of the estimator are shown in Fig. 7. The bias of the nonparametric estimator for cause 1 is inversely related to the percentile and increases in absolute value as the percentile increases. This feature can be attributed to the fact that the mean of the theoretical cause 1 failure times is larger than the mean of the theoretical cause 2 failure times. Therefore, typically observed values associated with cause 2, which are shorter, tend to censor the larger times generated from cause 1. To illustrate this, the average number of failure times observed for each cause were observed. The average number of type 1 failures observed for each iteration was 19.61 compared to the average number of type 2 failures observed for each iteration of 29.91.

7 Concluding remarks

A nonparametric resource-efficient method for estimating the system lifetime distribution with recurrent competing risks data was presented. The methods presented here extend the parametric model-based estimators in Taylor and Peña (2013), which utilized a recurrent structure for competing risks data via a perfect or partial repair strategy. In addition, this paper also implements an effective age process that allows for the modeling of various repair strategies. Under parametric modeling, methods for selecting the proper underlying distributions should be further explored. Alternatively, the nonparametric methods presented here offer flexibility in modeling, especially in scenarios where a parametric model may not be appropriate. The methods presented here were still limited in scope due to the assumed independence across and within systems. The independence assumption of component lifetimes within systems is somewhat restrictive and in realistic scenarios the inter-event times would most likely be correlated. Future work dealing with correlated component lifetimes will be needed. Most recently, Song and Kuo (2013) presented a dynamic subject-specific frailty model that allows for dependencies within subjects and with the dependencies evolving over time.

Acknowledgments

The authors would like to thank Professor Ananda Sen of the University of Michigan for providing the car warranty data. The authors thank the Reviewers, Associate Editor, and Editor for their comments that led to improvements. They are also grateful to Professors Jason Fine and Bo Lindqvist for inviting them to contribute to this special issue. Research partially supported by NSF Grant DMS1106435 and NIH Grants R01 CA154731 and P20 RR17698.

References

- Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, N. Springer series in statistics. New York: Springer; 1993. Statistical models based on counting processes.
- Bedford T, Lindqvist BH. The identifiability problem for repairable systems subject to competing risks. *Adv Appl Probab.* 2004; 36(3):774–790.

- Chen BE, Cook RJ. Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics*. 2004; 5:129–143. [PubMed: 14744832]
- Cook, RJ.; Lawless, JF. The statistical analysis of recurrent events. New York: Springer; 2007.
- Dauxois JY, Sencey S. Non-parametric tests for recurrent events under competing risks. *Scand J Stat*. 2009; 36(4):649–670.
- Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88(4):907–919.
- Geffray S. Modeling and inferential thoughts for consecutive gap times observed with death and censoring. *J Iran Stat Soc*. 2013; 12(1):71–112.
- Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics*. 2000; 56(2): 554–562. [PubMed: 10877316]
- Huang CY, Wang MC. Nonparametric estimation of the bivariate recurrence time distribution. *Biometrics*. 2005; 61(2):392–402. [PubMed: 16011685]
- Lawless J, Wigg M, Tuli S, Drake J, Lamberti-Pasculli M. Analysis of repeated failures of durations, with application to shunt failures for patients with paediatric hydrocephalus. *Appl Stat*. 2001; 50(4):449–465.
- Lindqvist BH. On the statistical modeling and analysis of repairable systems. *Stat Sci*. 2006; 21(4): 532–551.
- Peña, EA.; Strawderman, RL.; Hollander, M. Recent advances in reliability theory (Bordeaux, 2000. Birkhäuser Boston, Boston: Statistics for industry and technology; 2000. A weak convergence result relevant in recurrent and renewal models; p. 93-514.
- Peña EA, Strawderman RL, Hollander M. Nonparametric estimation with recurrent event data. *J Am Stat Assoc*. 2001; 96(456):1299–1315.
- Peña EA, Slate EH, González JR. Semiparametric inference for a general class of models for recurrent events. *J Stat Plan Inference*. 2007; 137(6):1727–1747. [PubMed: 19823592]
- Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981; 68(2):373–379.
- Song C, Kuo L. Dynamic frailty and change point models for recurrent events data. *J Iran Stat Soc*. 2013; 12(1):127–151.
- Taylor LL, Peña EA. Parametric estimation in a recurrent competing risks model. *J Iran Stat Soc*. 2013; 12(1):153–181.
- Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc*. 1989; 84(408):1065–1073.
- Xu J, Kalbfleisch JD, Tai B. Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*. 2010; 66(3):716–725. [PubMed: 19912171]

Recurrences of Failures

Fig. 1. Recurrences of two types of failure modes for 189 cars under warranty. Failure mode 1 (*blue triangle*) and failure mode 2 (*red circle*) (Color figure online)

Fig. 2. Nonparametric recurrent competing risks (nonparametric RCR), parametric recurrent competing risks (parametric RCR), and nonparametric non-recurrent competing risks estimates using the first event only (nonparametric CR) for the system life distribution are shown for the car warranty data (*top*). The nonparametric recurrent competing risks estimate with point-wise 95 % confidence bands are also shown (*bottom*) (Color figure online)

Randomly Generated Data for Two Weibull Causes

Fig. 3.

Randomly generated data using $\alpha_1 = 2$ and $\beta_1 = 1/2$ (green triangle) and $\alpha_2 = 3$ and $\beta_2 = 1/3$ (red inverted triangle) for 10 units with an exponential monitoring distribution with a mean of one (Color figure online)

Fig. 4.

The nonparametric recurrent competing risk estimate (*black, step*), parametric recurrent competing risk estimate (*blue, dashed*), and the theoretical (*red, solid*) cumulative hazard are displayed for cause 1 (*top*) and cause 2 (*bottom*) for the data in Fig. 3 (Color figure online)

System Lifetime Distribution

Fig. 5.

The nonparametric recurrent competing risk estimate (*black, solid*), estimate based on the misspecified parametric model (*green, dotted*), and the theoretical (*red, dashed*) system lifetime distribution are displayed for the data in Fig. 3 (Color figure online)

Fig. 6.

Plots of the means of the nonparametric recurrent competing risk estimates of the cumulative hazard (*black, solid*) for cause 1 (*top*) and cause 2 (*bottom*) over 1,000 simulations together with the theoretical cumulative hazard (*red, dashed*) are displayed for two Weibull competing risks (Color figure online)

Fig. 7.

The bias (*top row*) and mean squared error (*bottom row*) of the nonparametric recurrent competing risk estimator of the cumulative hazard for cause 1 (*first column*) and cause 2 (*second column*) are displayed for 1,000 iterations of two Weibull competing risks. The lines are plotted at every fifth percentile from the fifth percentile to the 95th percentile

Table 1

Summary statistics for the inter-event times associated with data displayed in Fig. 3

	Weibull (2,1/2)	Weibull (3,1/3)
Minimum	0.038	0.086
First quartile	0.287	0.216
Median	0.429	0.275
Mean	0.400	0.277
Third quartile	0.524	0.348
Maximum	0.792	0.436