Published in final edited form as:

Lifetime Data Anal. 2010 October ; 16(4): 509–524. doi:10.1007/s10985-009-9150-4.

Semiparametric analysis of recurrent events: artificial censoring, truncation, pairwise estimation and inference

Debashis Ghosh

Department of Statistics and Public Health Sciences, Penn State University, 514A Wartik Lab, University Park, PA 16802, USA

Debashis Ghosh: ghoshd@psu.edu

Abstract

The analysis of recurrent failure time data from longitudinal studies can be complicated by the presence of dependent censoring. There has been a substantive literature that has developed based on an artificial censoring device. We explore in this article the connection between this class of methods with truncated data structures. In addition, a new procedure is developed for estimation and inference in a joint model for recurrent events and dependent censoring. Estimation proceeds using a mixed U-statistic based estimating function approach. New resampling-based methods for variance estimation and model checking are also described. The methods are illustrated by application to data from an HIV clinical trial as with a limited simulation study.

Keywords

Accelerated failure time model; Cause-specific hazard; Comparability; Competing risks; Empirical process; Semi-competing risks data

1 Introduction

In many medical and scientific settings, analysts must deal with recurrent events data. Recurrent failure time data represents an event that can potentially repeat itself during the course of a study. Examples of recurrent events are tumor recurrences from oncology studies (Byar 1980), repeated cardiovascular events in patients (Cui et al. 2008) and births by a mother in demography studies. A problem arises when subjects who have repeated events tend to withdraw from the study and become censored observations. Such a dropout mechanism violates the usual independent censoring mechanism necessary for the validity of nonparametric and semiparametric inferences from recurrent failure time data (e.g., Andersen and Gill 1982; Pepe and Cai 1993; Lawless and Nadeau 1995).

To accommodate dependent censoring, a very popular approach has been to use inverse probability censoring weighted estimation techniques (Robins and Rotnitzky 1992). Such an approach requires that there exist a sufficiently rich set of covariates that fully explain the hazard of the dependent censoring mechanism. One then constructs a set of weights based on estimating this model that is then used to reweight censored observations in a manner that leads to asymptotically unbiased estimation. A generalization of this procedure is to construct so-called “doubly robust” estimating equations that yield consistent estimators if either the model for dependent censoring or for the event of interest is specified correctly. A summary of this

approach can be found in van der Laan and Robins (2002), and an application to the recurrent events problem has been given recently by Miloslavsky et al. (2004).

An alternative approach to this problem was initially proposed in the single-event case by Lin et al. (1996). They developed semiparametric estimation and inferential procedures for a bivariate accelerated failure time model for the time to event and for dependent censoring. While the model for the dependent censoring time can be estimated using standard techniques, that for the time to event cannot be handled in a straightforward way. This is because the dependent censoring induces a covariate-dependent dropout mechanism with respect to the time to event. Lin et al. (1996) introduce an artificial censoring device in order to construct an unbiased estimating function for the time to event. Recently, Peng and Fine (2007) developed a related estimation procedure for the same model. We will describe both of these approaches in Sect. 2.1.

For the situation of recurrent events, this approach has been extended by many authors. Ghosh and Lin (2003) developed an estimation procedure in which the target estimand was the expected number of events as a function of time. Such a model was extended to account for noncompliance in clinical trials by Matsui (2004). One could alternatively look at the times between the events, which are called interevent or gap times. For this approach, Chang (2000) proposed a scale-change model for analyzing the gap time between multiple events in the presence of dependent censoring. She also used the artificial censoring technique to construct an estimation procedure that would yield consistent estimators.

In this article, we revisit the model of Ghosh and Lin (2003) and use the ideas from Peng and Fine (2007) to construct an unbiased U-statistic-based estimating function. A secondary goal is provide a new perspective from which to interpret it. This is done using ideas from the analysis of truncated survival data (Bhattacharya et al. 1983; Wang et al. 1986). In particular, we shall see that an effect of the artificial censoring is to achieve comparability between pairs of observations, which will be shown in Sect. 2.2. A different interpretation of the artificial censoring was provided in a nice article by Joffe (2001), who used ideas from causal inference to explain the justification. The structure of this paper is as follows. In Sect. 2, we describe the observed data structure and review some previous methods for a univariate version of the problem (Lin et al. 1996; Peng and Fine 2007). It also provides an equivalence between the observed data structure with event times that are subject to right-truncation. We then describe the proposed estimation procedure and provide attendant asymptotic results in Sect. 3. A novel resampling-based method for variance estimation and goodness of fit techniques for model checking are also described. In Sect. 4, we describe application of the proposed methodology to data from an HIV clinical trial, as well as results from a limited simulation study. Finally, we conclude with some discussion in Sect. 5.

2 Preliminaries

2.1 Data and background

Define $a \wedge b$ to be the minimum of a and b . Let $N^*(t)$ denote the number of recurrent events that occur in $[0, t]$ in the absence of censoring and D denote the time to dependent censoring. Define C to be the time to independent censoring (i.e., time to random loss to follow-up and/or study termination) and Z to be a 0–1 treatment indicator. By convention, we will take 0 to represent placebo and 1 to represent treatment. We assume that $\{N^*(\cdot), D\}$ is independent of C given Z . Note, however, that we make no assumptions regarding the dependence between $N^*(\cdot)$ and D . As in Ghosh and Lin (2003), the observed data we consider are the following: $\{N_i(\cdot), X_i, \delta_i, Z_i\}$ ($i = 1, \dots, n$), a random sample from $\{N(\cdot), X, \delta, Z\}$, where $N(t) = N^*(t \wedge D \wedge C)$, $X = D \wedge C$, and $\delta = I(D \leq C)$.

The joint model we formulate is the following: $\{N_i^*(te^{\alpha_0 Z_i}), D_i e^{-\beta_0 Z_i}\}$ ($i=1, \dots, n$) have the same joint distribution. In the absence of dependent censoring, the model for the counting process is that given in Lin et al. (1998) and represents a natural extension of the AFT model to the situation of recurrent events.

Let us return to the case where N^* can take at most one jump, i.e., the event time of interest is univariate. This yields the following bivariate model:

$$\begin{bmatrix} T_i e^{-\alpha_0 Z_i} \\ D_i e^{-\beta_0 Z_i} \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} T_0 \\ D_0 \end{bmatrix} \quad (1)$$

in distribution, where T generically denotes the time to the event of interest. For model (1), estimation procedures have been proposed by Lin et al. (1996) and Peng and Fine (2007). We now review each of these in turn.

For the estimate of the regression coefficient in (1), we use the following log-rank estimating function:

$$\tilde{U}_1(\beta) = \sum_{i=1}^n \delta_i \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j(\beta) \geq \tilde{X}_i(\beta)\} Z_j}{\sum_{j=1}^n I\{\tilde{X}_j(\beta) \geq \tilde{X}_i(\beta)\}} \right], \quad (2)$$

where $\tilde{X}_i(\beta) = X_i \exp(-\beta' Z_i)$, $i = 1, \dots, n$. Let $\hat{\beta}$ be a zero-crossing of (2). This is the estimating function proposed by Louis (1981).

For the estimation of α , Lin et al. (1996) use an artificial censoring approach. The intuitive idea is to artificially trim the transformed time by a factor that allows for valid comparison between the two treatment groups. Define $\eta = (\alpha, \beta)$. This leads to the following estimating equation for estimation of η :

$$\tilde{U}_2(\eta) = \sum_{i=1}^n \tilde{\delta}_i^Y(\eta) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{Y}_j(\eta) \geq \tilde{Y}_i(\eta)\} Z_j}{\sum_{j=1}^n I\{\tilde{Y}_j(\eta) \geq \tilde{Y}_i(\eta)\}} \right], \quad (3)$$

$$\tilde{Y}_i(\eta) = \{T_i \exp(-\alpha Z_i) \wedge D_i \exp(-\beta Z_i - f) \wedge C_i \exp(-\beta Z_i - f)\},$$

where $\tilde{\delta}_i^Y(\eta) = I\{T_i \exp(-\alpha Z_i) \leq D_i \exp(-\beta Z_i - f) \wedge C_i \exp(-\beta Z_i - f)\}$

and $f = 0$ if $\alpha \leq \beta$ and $\beta - \alpha$ otherwise. Let $\hat{\alpha}$ be a zero-crossing of α from setting $\tilde{U}_2(\hat{\eta}) = 0$, where $\hat{\eta} = (\hat{\alpha}, \hat{\beta})$.

More recently, Peng and Fine (2007) introduced a modification to (3). Their idea is to perform the censoring on a pairwise basis between individuals. They propose the following data

transformation within pairs of observations i and j : $\{\tilde{Y}_{i(j)}(\eta), \tilde{\delta}_{i(j)}^Y(\eta), \tilde{Y}_{j(i)}(\eta), \tilde{\delta}_{j(i)}^Y(\eta)\}$, where

$$\begin{aligned} \tilde{Y}_{i(j)}(\eta) &= \{T_i \exp(-\alpha Z_i) \wedge D_i \exp(-\beta Z_i - g_{ij}) \wedge C_i \exp(-\beta Z_i - g_{ij})\}, \\ \tilde{\delta}_{i(j)}^Y &= I\{T_i \exp(-\alpha Z_i) \leq D_i \exp(-\beta Z_i - g_{ij}) \wedge C_i \exp(-\beta Z_i - g_{ij})\} \end{aligned}$$

and $g_{ij} = \max\{0, (\alpha - \beta)'Z_i, (\alpha - \beta)'Z_j\}$, where $1 \leq i, j \leq n$ and $\max$ denotes the maximum of these numbers. Note that as with f , we have suppressed dependence of g_{ij} on the regression parameters α and β . Using this transformation, Peng and Fine (2007) propose the following estimating function for η :

$$U_2^*(\eta) = 2n^{1/2} \times \sum_{i < j} \frac{(Z_i - Z_j) \left[\tilde{\delta}_{i(j)}^Y(\eta) I\{\tilde{Y}_{i(j)}(\eta) \leq \tilde{Y}_{j(i)}(\eta)\} - \tilde{\delta}_{j(i)}^Y(\eta) I\{\tilde{Y}_{j(i)}(\eta) \leq \tilde{Y}_{i(j)}(\eta)\} \right]}{n(n-1)}.$$

Peng and Fine (2007) argue that by considering pairwise censoring between individuals, this will lead to less artificial censoring than in the approach of Lin et al. (1996). They consider estimation of η using $\tilde{U}_1$ and U_2^* . Consistency and asymptotic normality of the resulting estimators is proven using results from U-statistic theory. Note that the estimating function proposed by Peng and Fine (2007) is in fact a U-statistic of order two. By contrast, the other estimating functions described so far have been U-statistics of order one. Note also that in the Peng and Fine approach, when Z is a binary covariate taking values zero and one, the only nonzero contributions to U_2^* occur when individuals i and j come from different treatment groups. Table 1 summarizes the values of $\tilde{Y}_{i(j)}(\beta)$ and $\delta_{i(j)}^Y(\beta)$.

2.2 Analogies with truncated data structures

For the moment, we consider a special case of (1) in which there is no censoring but that there is truncation so that (T, Z) is observed only if $T \leq t_0$, where t_0 is a truncation time. Then estimation relies on a notion of comparability between pairs of observations, which was introduced by Bhattacharya et al. (1983). To be specific, define $\Delta_{i(j)}(\alpha) = I\{T_i \exp(-\alpha Z_i) \leq t_0 \exp(-\alpha Z_j)\}$, $1 \leq i, j \leq n$. Then the pair of observations (i, j) are comparable if $\Delta_{i(j)} = 1$ for $i < j$. In particular, comparability pertains to the failure times (T_i, T_j) . Then following the construction of equation (2.1) in Bhattacharya et al. (1983), we can formulate the following class of estimating equations for α :

$$S(\alpha) = \sum_{i < j} W(Z_i, Z_j) \left[\Delta_{i(j)}(\alpha) I\{T_i \exp(-\alpha Z_i) \leq T_j \exp(-\alpha Z_j)\} - \Delta_{j(i)}(\alpha) I\{T_i \exp(-\alpha Z_j) \leq T_i \exp(-\alpha Z_i)\} \right],$$

where $W(x, y)$ represents a weight function; two choices for the weight function are $W_1(x, y) = 1$ and $W_2(x, y) = x - y$. Now we assume that there is no truncation but independent right-censoring by C . Then we can redefine the comparability indicator as

$\Delta_{i(j)}^C(\alpha) = I\{T_i \exp(-\alpha Z_i) \leq C_j \exp(-\alpha Z_j)\}$. Comparability of pairs of observations now pertain to the times $\{T_i, C_i\}$ and (T_j, C_j) . Note that $\Delta_{i(j)}^C$ is computable only if $I(T_i \leq C_i) = 1$, i.e., $T_i \leq C_i$. This leads to a modification of the previous estimating function:

$$S^C(\alpha) = \sum_{i < j} W^C(Z_i, Z_j) \left[I(T_i \leq C_i) \Delta_{i(j)}^C(\alpha) I\{T_i \exp(-\alpha Z_i) \leq T_j \exp(-\alpha Z_j)\} - I(T_j \leq C_j) \Delta_{j(i)}^C(\alpha) I\{T_j \exp(-\alpha Z_j) \leq T_i \exp(-\alpha Z_i)\} \right],$$

where W^C is weight function.

Now assume that there is dependent censoring by D in addition to independent censoring C . To define a notion of comparability, we require that the pairs (T_i, C_i, D_i) and (T_j, C_j, D_j) be comparable. Suppose that we transform all observed times by $\exp(-\alpha Z)$. Then this will induce differential censoring between the two treatment groups because all subjects in the group with

$Z = 1$ will have a censoring time $X \exp(-\alpha)$, while those in the control group (i.e., $Z = 0$) will have an independent censoring time X . This will lead to the groups being incomparable. To be precise, the failure times (T_i, C_i, D_i) and (T_j, C_j, D_j) will never be comparable when $Z_i \neq Z_j$. This is why one requires the censoring of X using $X \exp(-\beta Z)$, which leads to an unbiased comparison of cause-specific hazard functions for the time of interest. However, if $\alpha Z_i > \beta Z_i$, then a problem arises because then unobserved recurrent events are needed to facilitate the unbiased comparison. This is where the artificial censoring comes into play. Lin et al. (1996) and Ghosh and Lin (2003) propose censoring by $X \exp(-\beta Z - f)$. By contrast, Peng and Fine (2007) proposed using $X \exp(-\beta Z - g)$. Statistically, this will lead to fewer artificially uncensored observations. Conceptually, the artificial censoring done in the Peng-Fine approach is pairwise between individuals and ties in well with the notion of comparability, which compares observations in a pairwise manner.

3 Proposed methodology

3.1 Semiparametric estimation

Note that model of Ghosh and Lin (2003) implies that the event times corresponding to N^* satisfy the following model:

$$\log T_{ik} = \alpha_0 Z_i + \varepsilon, \quad (4)$$

where $N_i^*(t) = \sum_{k=1}^{\infty} I(T_{ik} \leq t)$, and ε is an error term. An equivalent formulation to the observed data process $N(\cdot)$ from Sect. 2.1. is to note that for the i th subject, we observe $\{T_{ij}, j = 1, \dots, K_i\}$ such that

$$T_{ij} \leq X_i, \quad (5)$$

where K_i is a discrete random variable taking nonnegative integer values. By definition, if $K_i = 0$, then we take the summation on the left-hand and right-hand side of (5) to be zero. We now condition on X and have the following:

Proposition—The data structure $\{N_i(\cdot), X_i, \delta_i, Z_i\} (i = 1, \dots, n)$ is equivalent to $(K_i, \{T_{ij}; j = 1, \dots, K_i\}, X_i, \delta_i, Z_i)$, where we observe all $T_{ij} \leq X_i$.

The proposition implies that we can treat X as a truncation time. Now suppose we want to estimate α_0 in the model (4); this is equivalent to the estimation in the joint model for $\{N^*(\cdot), T\}$ that we proposed in Sect. 1. Following the comparability ideas from the previous section, we develop an artificial censoring for each T_{ik} :

$$\tilde{T}_{i(j)k}(\eta) = T_{ik} \exp(-\alpha Z_i) \wedge (D_i \wedge C_i) \exp(-\beta Z_i - g_{ij}),$$

and

$$\tilde{\delta}_{i(j)k}(\eta) = I\{T_{ik} \exp(-\alpha Z_i) \leq (D_i \wedge C_i) \exp(-\beta Z_i - g_{ij})\}.$$

Based on this definition, we construct the following pairwise estimating function:

$$U_{ij}(\eta) = \sum_{k=1}^{K_i} \sum_{l=1}^{K_j} \times \left[\tilde{\delta}_{i(j)k}(\eta) I\{\tilde{T}_{i(j)k}(\eta) \leq \tilde{T}_{j(i)l}(\eta)\} - \tilde{\delta}_{j(i)l}(\eta) I\{\tilde{T}_{i(j)k}(\eta) \geq \tilde{T}_{j(i)l}(\eta)\} \right].$$

This leads to the following estimating function for α_0 :

$$U_2(\eta) = 2n^{1/2} \sum_{i < j} (Z_i - Z_j) \frac{U_{ij}(\eta)}{n(n-1)}. \quad (6)$$

Our procedure now is to solve for β using (2) as an estimating function for β ; denote the solution by $\hat{\beta}$. We then solve $U_2(\alpha, \hat{\beta}) = 0$ to obtain an estimator for α , $\hat{\alpha}$. Note that in the situation where K_i and K_j can take a maximum value of one, then (6) reduces to the estimating function considered in the univariate setting by Peng and Fine (2007).

Based on the joint model for recurrences and dependent censoring, setting $\eta = \eta_0$,

$$\begin{aligned} E\{U_2(\eta_0)\} &= \frac{1}{n(n-1)} \sum_{i < j} (Z_i - Z_j) E\{U_{ij}(\eta_0)\} \\ &= \frac{1}{n(n-1)} \sum_{i < j} (Z_i - Z_j) \times E\{E[U_{ij}(\eta_0) | \{\text{subjects } i \text{ and } j \text{ experience the } k\text{th event and are comparable}\}]\} = 0, \end{aligned}$$

where we have used the fact that for comparable events from comparable individuals, $P(\tilde{T}_{i(j)k}(\eta) \leq \tilde{T}_{j(i)l}(\eta)) = P(\tilde{T}_{i(j)k}(\eta) \leq \tilde{T}_{j(i)l}(\eta))$ for $\eta = \eta_0$. The conditioning event in the second equality has been written verbally so as to avoid further tedious notation. In other words, based on the comparability event, there is statistical independence of $\tilde{T}_{i(j)k}(\eta)$ and $\tilde{T}_{j(i)l}(\eta)$ for $\eta = \eta_0$. This principle also is present in Bhattacharya et al. (1983) and related works.

We can show $g_{ij} \leq f$ so that there is less artificial censoring with the proposed approach here relative to that of Ghosh and Lin (2003). However, this does not directly mean that the proposed method will necessarily always be more efficient. This is due to the fact that the proposed approach makes more comparisons (on the order of n^2) that are dependent relative to the approach of Ghosh and Lin (2003). When correlation between N^* and D is large then both approaches will discard a lot of data. While the estimation procedure will still be asymptotically unbiased, it will be at the cost of increased variance. One can in principle use the percentage of loss of case data an indication of strength of correlation between censoring and recurrent event measurements in general; formalizing this is beyond the scope of the current manuscript.

3.2 Asymptotic results

We now prove asymptotic properties about the estimators $\hat{\eta}$. Define

$$G(s, t, u, v, w) = Pr\left(\varepsilon_1^T - \varepsilon_1^D \leq s, \varepsilon_1^T - C_1 \leq t, \varepsilon_1^T - \varepsilon_2^T \leq u, \varepsilon_1^T - \varepsilon_2^D \leq v, \varepsilon_1^T - C_2 \leq w | Z_1, Z_2\right).$$

The following regularity conditions are imposed:

- A1. The parameter vector η_0 , the true value of η , is in an interior point in $\Theta \subset R^2$, where Θ is a compact parameter space.
- A2. $E[U_2(\alpha, \beta_0)] = 0$ has one unique solution at $\alpha = \alpha_0$.

A3. The partial derivatives of G are bounded, as are the marginal distributions of ε^T , ε^D , and C .

A4. $\text{Cov}\{(Z_1 - Z_2)k_1(Z_1, Z_2)^{1/2}\}$ and $\text{Cov}\{(Z_1 - Z_2)k_1(Z_1, Z_2)^{1/2}\}$ are positive definite, where

$$\begin{aligned} k_1(Z_1, Z_2) = & \frac{\partial G}{\partial s}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + \frac{\partial G}{\partial t}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + \frac{\partial G}{\partial v}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + \frac{\partial G}{\partial w}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + 2 \frac{\partial G}{\partial u}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \end{aligned}$$

and

$$\begin{aligned} k_2(Z_1, Z_2) = & \frac{\partial G}{\partial s}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + \frac{\partial G}{\partial t}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + \frac{\partial G}{\partial v}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \\ & + \frac{\partial G}{\partial w}(g_{12}, \\ & \exp(-\alpha_0 Z_1 - g_{12}), 0, g_{12}, \exp(-\alpha_0 Z_2 - g_{12})) \end{aligned}$$

We also assume the regularity conditions in Ying (1993). Based on these assumptions, we have the following result.

Theorem 1—Assuming conditions (A1)–(A4) and the conditions in Theorem 4(i) of Ying (1993), $\hat{\alpha}$ is a consistent estimator for α_0 .

Proof: In order to prove consistency, note that assumptions (A3) and (A4) imply that $v(\eta) \equiv E[n^{-1/2} U_2(\eta)]$ is differentiable at η_0 and that the partial derivatives of v with respect to α and β are nonsingular. In addition, we note that $\hat{\alpha}$ is bounded in probability. There exists a nonrandom function $m(\beta)$ such that

$$\sup_{\beta} \|n^{-1/2} U_1(\beta) - m(\beta)\| \rightarrow_p 0,$$

(Ying 1993). Condition (A2), combined with Ying (1993) assumption that β_0 is the only solution of $m(\beta) = 0$, implies that η_0 is the unique solution of $U_2\{m(\beta), v(\alpha, \beta)\} = 0$. To prove consistency of $\hat{\alpha}$, it suffices to show that

$$\sup_{\eta \in \Theta} \|n^{-1/2} U_2(\eta) - v(\eta)\| \rightarrow_p 0.$$

By the U-statistic law of large numbers, $\|n^{-1/2} U_2(\eta) - v(\eta)\| \rightarrow_p 0 \forall \eta \in \Theta$. Since Θ is compact, there exists a partition $\Theta_1, \dots, \Theta_m$ such that $\Theta \in \cup_{i=1}^m \Theta_i$ and where for $\eta^i \in \Theta_i$,

$$\max_{1 \leq i \leq m} \|n^{-1/2} U_2(\eta^i) - v(\eta^i)\| \rightarrow_p 0.$$

Applying arguments in the proof of Theorem 1 from (Peng and Fine (2007), Appendix), we have that uniform convergence follows because for every $\varepsilon > 0$, we can find a $\rho > 0$ such that

$$\lim_{n \rightarrow \infty} P \left(\sup_{\|\eta - \tilde{\eta}\| \leq \rho} n^{-1/2} \|U_2(\eta) - U_2(\tilde{\eta})\| \geq \varepsilon \right) = 0.$$

This proves the consistency of $\hat{\eta}$ for η_0 .

Theorem 2— $n^{1/2}(\hat{\alpha} - \eta_0)$ converges in distribution to a normal distribution with mean zero.

Proof: Using the arguments in Ying (1993), we can express U_1 as

$$U_1(\beta) = U_1(\beta_0) + n^{1/2} A(\beta - \beta_0) + o_p(1)$$

for $\beta: |\beta - \beta_0| \leq r$ for sufficiently small $r > 0$. For $U_2(\eta)$, we can apply assumptions (A2)–(A4), in conjunction with Lemma 2 of Honoré and Powell (1994) to imply that

$$\sup_{\eta \in N(\eta)} \frac{\|U_2(\eta) - U_2(\eta) - n^{1/2} v(\eta)\|}{1 + n^{1/2} \|v(\beta)\|} = o_p(1), \quad (7)$$

where $N(\eta)$ is a neighborhood of η_0 . Utilizing a Taylor series expansion of $v(\eta)$ about η_0 and combining with (7), we have that

$$\left\| v(\eta) - v(\eta_0) - \frac{\partial v(\eta_0)}{\partial \alpha}(\alpha - \alpha_0) - \frac{\partial v(\eta_0)}{\partial \beta}(\beta - \beta_0) \right\| = o(\|\eta - \eta_0\|).$$

This implies that

$$U_2(\eta) = U_2(\eta_0) + n^{1/2} [B_0 \ D_0](\eta - \eta_0) + o_p(1 + n^{1/2} \|\eta - \eta_0\|),$$

where $B_0 = \partial v(\eta_0)/\partial \alpha$ and $D_0 = \partial v(\eta_0)/\partial \beta$. Define $N_{1i}(t; \beta) = \delta_i I\{\tilde{X}_i(\beta) \leq t\}$ and

$$M_{1i}(t; \beta) = N_{1i}(t; \beta) - \int_0^t I\{\tilde{X}_i(\beta) \geq s\} \lambda_0(s) ds,$$

where $\lambda_0(t)$ is the hazard function of D_0 . Then

$$\begin{aligned} U_1(\beta) &= \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}^{(1)}(t; \beta)\} dN_{1i}(t; \beta) \\ &= \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}^{(1)}(t; \beta)\} dM_{1i}(t; \beta), \end{aligned}$$

where $\bar{Z}^{(1)}(t; \beta) = \sum_{j=1}^n I\{\tilde{X}_j(\beta) \geq t\} Z_j / \sum_{j=1}^n I\{\tilde{X}_j(\beta) \geq t\}$. Because $M_{1i}(t) \equiv M_{1i}(t; \beta_0)$ ($i = 1, \dots, n$) are orthogonal martingales with respect to the marginal filtration $\mathcal{F}_t = \sigma\{N_{1i}(s; \beta_0), I\{\tilde{X}_i(\beta_0) \geq s\}, \mathbf{Z}_i; 0 \leq s \leq t, i = 1, \dots, n\}$,

$$n^{-1/2} U_1(\beta_0) = n^{-1/2} \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{z}^{(1)}(t)\} dM_{1i}(t) + o_p(1), \quad (8)$$

where $\bar{z}^{(1)}(t)$ is the probability limit of $\bar{Z}^{(1)}(t; \beta_0)$. Similarly, for $U_2(\eta)$,

$$U_2(\eta_0) = \sum_{i=1}^n 2h_1(\mathbf{V}_i, \eta_0) + o_p(1), \quad (9)$$

where $\mathbf{V}_i = (\varepsilon_{i1}^T, \dots, \varepsilon_{ik_i}^T, D_i, C_i, Z_i)$ ($i = 1, \dots, n$), and $h_1(\mathbf{v}, \eta) = E[h(\mathbf{v}, \mathbf{V}_2, \eta)]$, and $h(\mathbf{V}_1, \mathbf{V}_2, \eta) = (Z_1 - Z_2)U_{ij}(\eta)$. We can thus combine representations (8) and (9) into a representation of $n^{-1/2} \mathbf{U}(\eta_0) \equiv n^{-1/2} \{U_1(\beta_0), U_2(\eta_0)\}$ as a normalized average of independently and identically distributed random vectors with zero mean. The asymptotic normality then follows by the multivariate central limit theorem. One can then perform a multivariate Taylor series expansion followed by appealing to Slutsky's theorem to show that $n^{1/2}(\hat{\eta} - \eta_0)$ is bivariate normal with a zero-mean vector and variance-covariance matrix $\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$, where

$$\mathbf{A} = \begin{pmatrix} \partial U_1(\alpha_0)/\partial \alpha & 0 \\ \partial U_2(\eta)/\partial \alpha & \partial U_2(\eta)/\partial \beta \end{pmatrix}$$

and $\mathbf{B}$ is the limiting variance-covariance matrix of $n^{-1}\mathbf{U}(\eta_0)$.

3.3 Variance estimation and model checking

Variance estimation for $\hat{\eta}$ based on the result of Theorem 2 is problematic in that it requires estimation of the derivative of the density of the error distributions. We take an alternative approach based on resampling recently proposed by Zeng and Lin (2008). The algorithm works as follows:

1. Generate observations $\mathbf{G} \equiv (G_1, G_2)$, two independent and identically distributed $N(0, 1)$ random variables.
2. Calculate $n^{-1/2}[U_1(\hat{\beta} + n^{-1/2}G_1), U_2(\hat{\alpha} + n^{-1/2}\mathbf{G})]$.
3. Repeat steps 1 and 2 M times.
4. Regress $n^{-1/2}U_1(\hat{\beta} + n^{-1/2}G_1)$ on G_1 across the M datasets. Similarly, Regress $n^{-1/2}U_2(\hat{\alpha} + n^{-1/2}\mathbf{G})$ on G_1 and G_2 across the M datasets.
5. Estimate the variance-covariance matrix of $n^{1/2}(\hat{\eta} - \eta_0)$ as $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$, where the first row of $\hat{\mathbf{A}}$ is the slope estimate from the first regression in the previous step, while the second row of $\hat{\mathbf{A}}$ are slope estimates from the second regression in the previous step.

Modifying the arguments in Zeng and Lin (2008), it can be shown that this algorithm provides a consistent estimator of the variance-covariance matrix of $n^{1/2}(\hat{\eta} - \eta_0)$. The main condition needed for validity is that one be able to take a Taylor series expansion around η_0 ; such a claim is given implicitly in the proof of Theorem 2. Note that this proposed resampling scheme is much faster than that proposed in Ghosh and Lin (2003). In particular, this algorithm avoids having to solve the estimating equation for the perturbed datasets. By contrast, the method of variance estimation in Ghosh and Lin (2003) does involve solving the estimating functions for α and β repeatedly. In simulation studies not shown here, we found the proposed method performed similarly in finite samples compared to the original approach of Ghosh and Lin (2003), but the computational speed of the new method is substantially faster. Zeng and Lin (2008) found in their experience that the method could be 100–1000 times faster than resampling methods of the type proposed by Ghosh and Lin (2003).

In practice, it will be important to check the joint modelling assumption. Based on the notation of Lin et al. (1994), we suggest checking the model by using the following statistics: $\sup_t \|U_1(t; \hat{\beta})\|$ and $\sup_t \|U_2(t; \hat{\eta})\|$. These are referred to as the score processes for dependent censoring and recurrences, respectively. The corresponding processes based on the observed data are given by $U_1(t; \hat{\beta})$ and $U_2(t; \hat{\eta})$. If the joint model holds, then these processes will fluctuate randomly about zero. We construct test statistics based on the joint distribution of $\mathbf{U}(s, t; \hat{\eta}) = \{U_1(s; \hat{\beta}), U_2(t; \hat{\eta})\}'$ and obtain the null distribution using the nonparametric bootstrap, as in Peng and Fine (2007).

4 Numerical examples

4.1 HIV clinical trial

HIV infection in humans is now pandemic. As of January 2006, the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) estimate that AIDS has killed more than 25 million people since it was first recognized on December

1, 1981; it is estimated that 0.6% of the world's population is infected with HIV (UN 2006). HIV is an infection that attacks the white blood cells in a person's body. This causes the immune system to fail, leading to the development of opportunistic infections that are life-threatening. It is therefore paramount to find treatments that can retard the progression of this disease and ultimately prolong survival.

We apply the proposed methodology to data from a trial conducted by Abrams et al. (1994). They compared two treatments for HIV, didanosine (ddI) and zalcitabine (ddC). In the study, 467 patients who were intolerant of or had failed treatment with zidovudine (ZDV) were randomized to ddI or ddC. There were 237 subjects in the ddC ($Z = 1$) group and 230 in the ddI ($Z = 0$) arm. Abrams et al. (1994) considered three endpoints: survival, time to disease progression, and time to disease progression or death. Disease progression was defined as the first occurrence of an AIDS-defining condition as described in a 1987 Centers for Disease Control classification or as a recurrence of the following opportunistic conditions: *Pneumocystis carinii* (*P. carinii*) pneumonia, esophageal candidiasis, herpes simplex infection, disseminated herpes zoster, and septicemia due to nontyphoidal salmonella. Table 2 reports the number of progression events experienced by treatment group. Note that the progression event can recur so that the methodology developed in this paper is applicable. Figure 1 plots the Kaplan-Meier estimators for time to death in the two treatment arms. Based on the graph, we see that while there appears to be no difference in survival for the first six months, there is evidence for survival benefit on ddC relative to ddI later in the trial.

We first apply the methodology of Ghosh and Lin (2003). For the estimate of treatment on death, the estimated effect is 0.17 with an estimated standard error of 0.21, while that on the recurrences is -0.05 with an estimated standard error of 0.32. For variance estimation of the treatment effects, we adapted the methodology of Zeng and Lin (2008), described for the current methodology in Sect. 3.3., instead of the resampling scheme utilized by Ghosh and Lin (2003). All results here use $M = 10,000$ realizations. Thus, while subjects on ddC tend to live longer than those on ddI, their time between progression events is shorter. However, neither effect is statistically significant at a significance level of 0.05. For this approach, 9.3% of the observations were artificially censored. We now use the proposed methodology, which has the same estimated effect of treatment on death, but now the treatment effect on recurrence is -0.03 with an estimated standard error of 0.28. While we get different estimates for the treatment effect on recurrences and its associated standard error, qualitatively the finding is in agreement with the first analysis. In the proposed method, 7.8% of the observations were artificially censored.

Plots of the goodness of fit statistics from Sect. 3.3 are given in Figs. 2 and 3. The p -values, based on 1000 simulations, for assessing goodness of fit for survival and recurrences are 0.28 and 0.48. This suggests that there is not sufficient evidence to reject the joint model fitted here.

4.2 Simulation studies

We compared the proposed methodology to that of Ghosh and Lin (2003) in a limited simulation study. The bivariate vector $\{N_0^*(t), D_0\}$ was simulated using a two-stage procedure. A gamma random effect v , with mean 1 and variance σ^2 , was first generated. Conditional on v , the gap times corresponding to $N_0^*(\cdot)$ and the time to dependent censoring (D_0) were generated as independent exponential random variables with rates $4v^{-1}$ and v^{-1} . It can be shown that such a probabilistic specification is consistent with the joint model in this paper.

If σ^2 is nonzero, then induce correlation between recurrences and dependent censoring. For each simulation study, we considered $\sigma^2 = 0, 1.0$ and 4.0 . We compared the estimators from the proposed method to those using the Ghosh and Lin (2003) method. For each simulation

setting, 1000 simulation samples were considered, and 1000 resamplings were generated for each simulation sample. We took $\eta_0 = [0.7, 1.2]'$; the results only deal with the finite-sample properties of estimation for α_0 . Independent censoring is distributed as a Uniform[0, 10] random variable, which yielded an average of approximately 2.8 observed recurrences. We see from Table 3 that while both the proposed methods and that of Ghosh and Lin (2003) provide satisfactory finite-sample behavior, there is approximately 50% less artificial censoring performed on average with the new method, which leads to efficiency gains in the estimation of α .

5 Discussion

In this article, we have developed a new approach to estimation in the joint model considered by Ghosh and Lin (2003). Here, a pairwise approach to artificial censoring was used; the work presented here can be viewed as a recurrent events analog of the work of Peng and Fine (2007). The simulation study suggests that the pairwise censoring can lead to gains in efficiency of estimating the treatment effect relative to the work of Ghosh and Lin (2003). However, the parameter that influences the situation is the correlation between recurrences and dependent censoring. If there is a lot of artificial censoring, it is preferable to attempt to use the approach of van der Laan and Robins (2002), possibly combined with some type of sensitivity analysis.

Ideally, it would be desirable to incorporate noncompliance into this framework, as did Matsui (2004). One point to note is that for that method to work, the independent censoring time must be known for all individuals in the study. This is feasible in a clinical trial setting where the dates for study termination and patient randomization are known for all individuals; the independent censoring time is nothing more than the difference between the two. If these dates are not available, then C_0 will not be known. In the clinical trial example considered by Matsui (2004), C_0 was not known, so several hypothetical values of C_0 were used. However, the method of Matsui does not seem to extend easily into the pairwise-dependent artificial censoring technique used here. It is still not clear how to allow for noncompliance in the recurrent event setting considered here.

Finally, a secondary theme of the work presented here is to show that there are parallels in the data structure considered here and data subjected to right-truncation. For truncated data, the notion of comparability of event times is important. We have shown in Sect. 2.2. how to generalize this notion to the dependent censoring. Further exploration should allow for development of new estimation procedures for non-AFT models. This research is currently under investigation.

While we have presented a limited comparison of the proposed method with that of Ghosh and Lin (2003), further investigations are needed in order to recommend one procedure over another in practice. As discussed in Sect. 2, one procedure does not dominate the other in terms of estimation efficiency because of the number of comparisons that are used in each procedure. More research is needed to determine the suitability of the two procedures for real data analyses.

Acknowledgments

The author would like to acknowledge the support of National Institutes of Health grant CA129102. He thanks Jim Neaton for the use of the data from the HIV clinical trial.

References

- Abrams DI, Goldman AI, Launer C, Korvick JA, Neaton JD, Crane LR, Grodesky M, Wakefield S, Muth K, Kornegay S, et al. A comparative trial of didanosine or zalcitabine after treatment with zidovudine

- in patients with human immunodeficiency virus infection. The Terry Bein community programs for clinical research on AIDS. *N Engl J Med* 1994;330:657–662. [PubMed: 7906384]
- Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat* 1982;10:1100–1120.
- Bhattacharya PK, Chernoff H, Yang SS. Nonparametric estimation of the slope of a truncated regression. *Ann Stat* 1983;11:505–514.
- Byar, DP. The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine and topical thiotepa. In: Pavone-Macaluso, M.; Smith, PH.; Edsmyr, F., editors. *Bladder tumors and other topics in urological oncology*. Plenum; New York: 1980. p. 363-370.
- Chang SH. A two-sample comparison for multiple ordered event data. *Biometrics* 2000;56:183–189. [PubMed: 10783794]
- Cui J, Forbes A, Kirby A, Marschner I, Simes J, West M, Tonkin A. Parametric conditional frailty models for recurrent cardiovascular events in the lipid study. *Clin Trials* 2008;5:565. [PubMed: 19029205]
- Ghosh D, Lin DY. Semiparametric analysis of recurrent events in the presence of dependent censoring. *Biometrics* 2003;59:877–885. [PubMed: 14969466]
- Honoré B, Powell JL. Pairwise difference estimators of censored and truncated regression models. *J Econom* 1994;64:241–278.
- Joffe MM. Administrative and artificial censoring in censored regression models. *Stat Med* 2001;20:2287–2304. [PubMed: 11468764]
- Joint United Nations Programme on HIV/AIDS. Overview of the global AIDS epidemic. 2006 [Accessed September 18, 2008]. Available at http://data.unaids.org/pub/GlobalReport/2006/2006_GR_CH02_en.pdf
- Lawless JF, Nadeau JC. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995;37:158–168.
- Lin DY, Fleming TR, Wei LJ. Confidence bands for survival curves under the proportional hazards model. *Biometrika* 1994;81:73–81.
- Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* 1996;83:381–393.
- Lin DY, Wei LJ, Ying Z. Accelerated failure time models for counting processes. *Biometrika* 1998;85:605–618.
- Louis TA. Nonparametric analysis of an accelerated failure time model. *Biometrika* 1981;68:381–390.
- Matsui S. Analysis of times to repeated events in two-arm randomized trials with noncompliance and dependent censoring. *Biometrics* 2004;60:965–976. [PubMed: 15606417]
- Miloslavsky M, Keles S, van der Laan MJ, Butler S. Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J R Stat Soc Ser B* 2004;66:239–257.
- Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. *J Am Stat Assoc* 2007;101:1085–1093.
- Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J Am Stat Assoc* 1993;88:811–820.
- Robins, J.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, N.; Dietz, K.; Farewell, V., editors. *Aids epidemiology-methodological issues*. Birkhauser; Boston: 1992. p. 297-331.
- van der Laan, MJ.; Robins, JM. *Unified methods for censored longitudinal data and causality*. Springer; New York: 2002.
- Wang MC, Jewell NP, Tsai WY. Asymptotic properties of the product limit estimate under random truncation. *Ann Stat* 1986;14:1597–1605.
- Ying Z. A large-sample study of rank estimation for censored regression data. *Ann Stat* 1993;21:76–99.
- Zeng D, Lin DY. Efficient resampling methods for nonsmooth estimating functions. *Biostatistics* 2008;9:355–363. [PubMed: 17925303]

Fig. 1. Kaplan–Meier (K–M) estimators for time to death in the ddI (*yellow line*) and ddC (*blue line*) groups in the CPCRA study. A log-rank test for time to death between the two treatment arms yielded a test statistic of 2.1, which corresponds a p -value of 0.15. In the print version of this figure, the K–M estimator for the ddI group is lower than that of the ddC group at day 300, but this ordering is reversed at day 575

Fig. 2. Cumulative martingale sum process for treatment effect on death for the observed data (*solid line*) and 10 realizations from simulated null distribution (*dashed lines*)

Fig. 3. Cumulative martingale sum process for treatment effect on recurrences for the observed data (*solid line*) and 10 realizations from simulated null distribution (*dashed lines*)

Table 1

Values of $\tilde{Y}_{(i)j}(\eta)$ and $\tilde{\delta}_{i(j)}^Y$ from the method of Peng and Fine (2007)

Z_i	Z_j	$\alpha \leq \beta$	$\tilde{Y}_{i(j)}(\eta)$	$\tilde{\delta}_{i(j)}^Y$
0	1	True	$T_i \wedge D_i \wedge C_i$	$I(T_i \leq D_i \wedge C_i)$
0	1	False	$T_i \wedge (D_i \wedge C_i) \exp\{-(\alpha - \beta)\}$	$I(T_i \leq (D_i \wedge C_i) \exp\{-(\alpha - \beta)\})$
1	0	True	$T_i \exp(-\alpha) \wedge (D_i \wedge C_i) \exp(-\beta)$	$I(T_i \exp(-\alpha) \leq (D_i \wedge C_i) \exp(-\beta))$
1	0	False	$(T_i \wedge D_i \wedge C_i) \exp(-\alpha)$	$I(T_i \exp(-\alpha) \leq (D_i \wedge C_i) \exp(-\alpha))$

Table 2

Opportunistic infections in HIV trial

Treatment	Subjects	Number of opportunistic infections					
		0	1	2	3	4	5
ddl	230	110	76	37	6	1	0
ddC	237	122	66	29	14	5	1

Table 3

Summary of simulation results for α_0

n	σ^2	Ghosh and Lin (2003) method					Proposed method				
		Bias	SE	SEE	CP	ACP	Bias	SE	SEE	CP	ACP
100	0	-0.01	0.12	0.11	0.94	0.25	-0.01	0.10	0.09	0.95	0.12
	1	0.01	0.18	0.19	0.95	0.29	0.01	0.13	0.11	0.94	0.15
	4	-0.02	0.21	0.20	0.94	0.31	0.01	0.17	0.17	0.95	0.19

Note: Bias is the mean of the estimators of α_0 minus α_0 ; SE is the standard error of the estimators of α_0 ; SEE is the mean of the standard error estimate; CP is the coverage probability of the Wald 95% confidence interval, and ACP is the artificial censoring proportion. Independent censoring was generated using a Uniform (0, 10) random variable.