Published in final edited form as:
Journal of the American Statistical Association (2009), **104**(485): 60–75
DOI: 10.1198/jasa.2009.0004

Robust Estimation of Mean Functions and Treatment Effects for Recurrent Event Under Event-Dependent Censoring and Termination: Application to Skeletal Complications in Cancer Metastatic to Bone

RICHARD J COOK

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: rjcook@uwaterloo.ca

JERALD F LAWLESS

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: jlawless@uwaterloo.ca

LAJMI LAKHAL-CHAIEB

Department of Mathematics and Statistics,
University of Laval, Quebec City, QC, G1K 7P4, Canada
E-mail: lajmi.lakhal@mat.ulaval.ca

KER-AI LEE

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada

Summary

In clinical trials featuring recurrent clinical events, the definition and estimation of treatment effects involves a number of interesting issues, especially when loss to follow-up may be event-related and when terminal events such as death preclude the occurrence of further events. This paper discusses a clinical trial of breast cancer patients with bone metastases where the recurrent events are skeletal complications, and where patients may die during the trial. We argue that treatment effects should be based on marginal rate and mean functions. When recurrent event data are subject to event-dependent censoring, however, ordinary marginal methods may yield inconsistent estimates. Incorporating correctly specified inverse probability of censoring weights into analyses can protect against dependent censoring and yield consistent estimates of marginal features. An alternative approach is to obtain estimates of rate and mean functions from models that involve some conditioning to render censoring conditionally independent. We consider three methods of estimating mean functions of recurrent event processes and examine the bias and efficiency of unweighted and inverse probability weighted versions of the methods with and without a terminating event. We compare the methods via simulation and use them to analyse the data from the breast cancer trial.

Keywords: Dependent censoring; Inverse probability of censoring weights; Recurrent event; Terminal event

1 INTRODUCTION

1.1 SKELETAL COMPLICATIONS IN PATIENTS WITH BONE METASTASES

The motivating study for the present article is a randomized clinical trial designed to investigate the effect of a biosphosphonate, pamidronate, on the development of skeletal complications (e.g., fractures, spinal cord compression) in breast cancer patients with bone metastases (Hortobagyi et al., 1996). Patients were accrued to this study between January 1991 and March 1994 from 97 sites in the United States, Canada, Australia and New Zealand. Patients with stage IV breast cancer receiving cytotoxic chemotherapy with at least one predominantly lytic bone lesion greater than or equal to one centimeter in diameter were randomized within strata defined by ECOG status.

A total of 382 women were enrolled in the study with 185 randomized to receive pamidronate and 197 to placebo control. Two patients randomized to placebo did not have bone metastases and were therefore excluded from subsequent analyses. Patients randomized to the pamidronate arm received 90 mg of pamidronate disodium via a two hour infusion every four weeks, whereas patients randomized to the placebo received dextrose infusions; however those on a three week chemotherapy regimen were permitted to receive the study drug every three weeks. At monthly visits patients were assessed and the occurrence of skeletal complications was recorded. The skeletal complications of interest include pathologic fractures, spinal cord compression with vertebral fracture, the need for surgery to treat or prevent fractures, and the need for radiation for the treatment of bone pain. After completion of the planned one year follow-up, the observation was extended for an additional year and the results were published in Hortobagyi et al. (1998). Each patient was followed until death, the last date of contact or loss to follow-up, or February 1, 1996.

Table 1 contains the frequencies of the numbers of skeletal complications experienced by the patients over the course of observation. There is a higher proportion of patients having no skeletal complications in the pamidronate arm compared with the placebo arm, but crude summaries of this sort are difficult to interpret, because death is a terminal event precluding the occurrence of further skeletal complications, and patients are withdrawn from this study or may be lost to follow-up. Figure 1 contains the Kaplan-Meier estimates of the cumulative probability of death for placebo and pamidronate treated patients, which reveals comparable risk of death.

It is common in settings like this to see analyses based on the complication-free survival time (that is, time to the first event). With this endpoint, standard methods of survival analysis can be used including Kaplan-Meier plots, Cox regression, and log-rank tests for treatment effects (Kalbfleisch and Prentice, 2002; Lawless, 2003). However, if a palliative therapy delays the occurrence of skeletal complications (the clinical event of interest), but does not affect the mortality rate, this may underestimate the effect of treatment on the skeletal event of interest, particularly if many patients die without experiencing this event. Complication-free survival analyses are often motivated by the fact that they obviate the need for dealing with the competing risk problem arising from deaths. A competing risk analysis (Crowder, 2001) of the time to the first skeletal event is often preferable, however, because it is directed at the event of primary interest; standard Cox regression models can be fit for the time to first skeletal event, with inferences based on cause-specific hazards and regression coefficients.

A shortcoming of analyses based on the time to first skeletal complication is that they do not make use of the information on complications arising after the first one. Use of subsequent clinical events is desirable when more comprehensive summaries of the disease burden are of interest or when cost implications for the nonfatal events are of interest in an economic analysis of treatment. Simple

Table 1: Frequency distribution for the total number of skeletal complications observed by treatment arm in Hortobagyi et al. (1998)

Number of	Placebo		Pamidronate	
Skeletal Events	No.	Percent	No.	Percent
0	69	35.4	99	53.5
1	41	21.0	39	21.1
2	34	17.4	17	9.2
3	18	9.2	13	7.0
4	12	6.2	10	5.4
5	7	3.6	3	1.6
6	5	2.6	1	0.5
7	4	2.1	1	0.5
$\geq$8	5	2.6	2	1.1
Total	195		185	

summaries of the total number of complications observed on study do not address the variation in censoring and survival times and as a result are uninterpretable. Summaries based on events-per-person-years statistics are often provided (e.g., Kanis et al., 1996), but these too are problematic because they typically do not distinguish between the end of observation due to censoring and death; for the former the event process may continue but for the latter it is terminated. Moreover, analyses must deal with possible association between the recurrent event process and the terminating death process.

In the context of skeletal complications, Moecks et al. (2004) review a variety of methods with respect to clinical utility and statistical validity. Crude data summaries based on simple or period-specific event rates are considered, and formal comparisons are recommended based on nonparametric tests. Such approaches do not require modeling assumptions, but given the possibly complex relationship between recurrent events and death, they are not based on interpretable summary statistics. The present article takes a rigorous look at the specification and estimation of treatment effects in this setting. Finally, as we will find in the bone metastases study, there may even be event-related loss to follow-up of persons still alive. In this case adjustments are needed to methods mentioned previously to avoid bias in estimates.

1.2 Frameworks for Assessing Treatment Effects

There have been considerable advances in the past few decades on statistical methods for the analysis of recurrent events; see Andersen et al. (1993) and Cook and Lawless (2007) for comprehensive reviews. These advances have been motivated, in part, by the many settings in medical research in which recurrent events arise. Patients with asthma, for example, may have repeated attacks (Duchateau et al., 2003), persons with cystic fibrosis suffer repeated pulmonary infections (Therneau and Hamilton, 1997), and persons with epilepsy can experience repeated seizures (Thall and Vail, 1990).

Intensity-based methods, where the probability of the occurrence of a new event is modeled in terms of previous event history, are appealing in many contexts and offer a wide range of analyses. However, they are rarely used as a basis for examining treatment effects in clinical trials for a number of reasons. First, the effects of treatment or other baseline covariates should be easily interpreted and understood in clinical trials, and should provide a convenient basis for trial design. The fact that

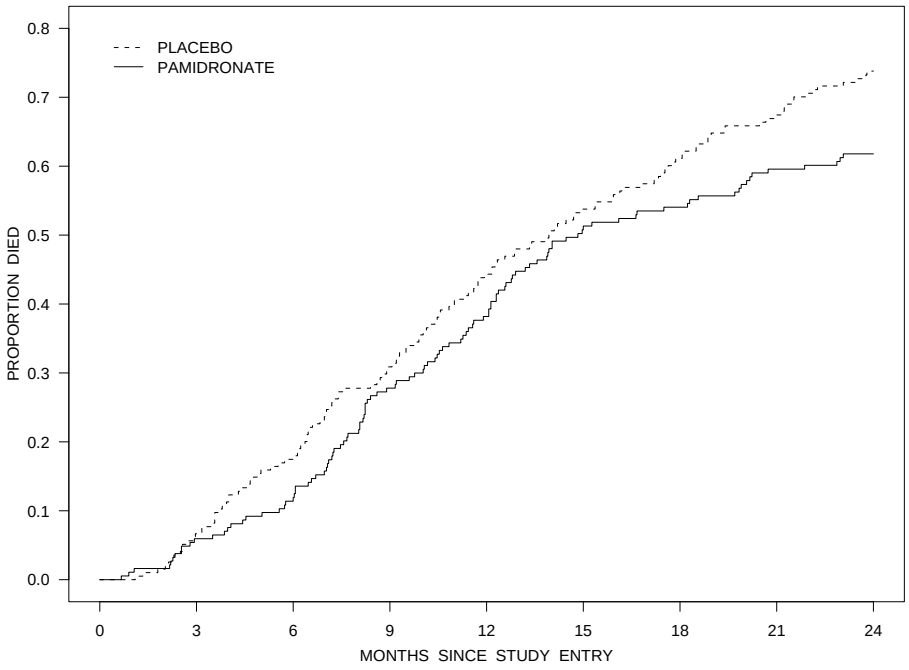

Figure 1: Kaplan-Meier estimate of survival distribution for placebo and pamidronate treated patients in Hortobagyi et al. (1998).

intensity-based methods in effect condition on dynamic internal covariates (i.e., prior event history) makes interpretation of treatment effects difficult, and in particular confounds treatment effect with prior event history. In addition, it is generally important that the comparison of treatments can be based on their random assignment to subjects. Conditioning on prior event history at any given time does not permit this, because balance across other factors that is achieved by randomization is typically destroyed. Finally, intensity-based methods are susceptible to model misspecification.

Over the past 20 years there has been considerable development of methods based on marginal features of recurrent event processes, and these provide important tools for the design and analysis of clinical trials. The marginal features that have been proposed most often for analysis are (1) event rate functions or cumulative mean functions (e.g., Cox and Lewis, 1966; Andersen and Gill, 1982; Lawless, 1987a,b; Pepe and Cai, 1993; Lawless and Nadeau, 1995; Lawless et al., 1997), (2) times to specific events (e.g., Wei et al., 1989; Wei and Glidden, 1997), and (3) times between successive events (Lin et al., 1999; Schaubel and Cai, 2004). As discussed by Cook and Lawless (2007, Sec. 8.4), the use of rate and mean functions is the most broadly appealing and applicable approach, and we focus on this here.

Rate and mean functions are defined as follows. For simplicity, consider a single homogeneous group of subjects, and let $N_i(t)$ denote the number of events experienced by subject i over the time period $(0, t]$. Then the event mean and rate functions are, respectively,

$$\mu(t) = E\{N_i(t)\} \qquad \rho(t) = \mu'(t),$$

where we assume that events occur in continuous time. To compare two treatments, denoted 1 and 2, we can conveniently consider ratios $\mu_1(t)/\mu_2(t)$ or $\rho_1(t)/\rho_2(t)$, or differences $\mu_1(t) - \mu_2(t)$ or $\rho_1(t) - \rho_2(t)$. In many settings it is found that the rate ratio $\rho_1(t)/\rho_2(t)$ is more or less a constant θ, and in that case $\mu_1(t)/\mu_2(t)$ is also approximately equal to θ. Methods of analysis are well developed

in this case, provided follow-up is independent of event occurrence and that there are no terminating events (Cook and Lawless, 2007, Chap. 3).

When the recurrent events arise in populations with advanced disease, mortality rates can be high and generalizations are required to distinguish between patients who are lost to follow-up and those whose follow-up is terminated by death. In other settings there may be events that terminate recurrent event processes. For example, when the recurrent events are because of a transient condition that can be cured by treatment, they will cease to occur when the underlying condition is resolved. Some authors have recommended analyses based on shared or bivariate frailty models that link the recurrent event and mortality processes (Huang and Wang, 2004; Liu et al., 2004; Rondeau et al., 2007). Whereas this approach relates recurrent events and death, the event intensity of this model is nonzero after death and so it does not correspond to any plausible physical process. In addition, marginal rate, mean and survivor functions are complicated functions of treatment. Cumulative mean functions that account for the possible termination of the event process can be defined in settings with terminal events. Cook and Lawless (1997) describe a simple method for estimation of a marginal cumulative mean function and Ghosh and Lin (2000, 2002) extend this and develop robust two-sample tests based on marginal mean functions and survival distributions.

Many marginal methods are robust to the type of underlying recurrent event process, but are valid only under independent censoring. Therefore, unlike intensity-based analyses that are valid under adaptive (conditionally independent) censoring, biased estimates of rate functions and other marginal features are obtained when subjects are withdrawn from a study based on their prior response to treatment. Several authors have proposed the use of estimating equations with weights that are inversely proportional to the probability of censoring (Robins and Rotnitzky, 1992, 1995; Robins, 1993, Satten et al., 2001; Datta and Satten, 2002). Such approaches require one to model the censoring process, but when this is done successfully consistent estimates of marginal features are obtained under event-dependent censoring.

An alternative approach is to develop methods that condition on enough of the event history to render the censoring conditionally independent, and then to obtain marginal features by averaging over prior event history. This is often complex, but in some cases can be achieved with an approach based on multistate Markov models, which have recently been found to give robust estimates of state occupancy probabilities (Aalen et al., 2001; Datta and Satten, 2001; Gunnes et al., 2007). This robustness makes this an appealing approach because, in addition to providing an estimate of the mean function, it allows for robust estimation of the marginal distribution of the number of cumulative number of events over time. Moreover it is easily adapted to deal with the presence of terminal events. The inverse probability of censoring weights described by Datta and Satten (2002) can also be applied if there is concern about residual dependence between the censoring and event processes.

The purpose of this article is to study different methods for estimation of event mean functions, with specific attention to the assessment of treatment effects in clinical trials. We consider methods based on marginal rate functions, marginal survivor functions for event times, and partially conditional rate functions employing Markov assumptions. Inverse probability of censoring weighted (IPCW) versions of the estimators are also given to provide robustness to event-dependent censoring. We examine the performance of these methods for different types of event processes under independent and event-dependent censoring. Extensions that deal with a terminating event are also presented and investigated. The methods are then used to analyse the data from the motivating trial of breast cancer patients with skeletal metastases, and to make treatment comparisons.

The remainder of the article is organized as follows. In Section 2 we introduce some notation and briefly review intensity-based modeling, marginal features, and censoring of recurrent event processes. Nonparametric methods for estimating mean functions for recurrent events are given in Section 3, and extensions are given in Section 4 for recurrent and terminal events. The design and results of simulation studies are described in Section 5 and in Section 6 the nonparametric methods

are applied to the study of skeletal complication in patients with bone metastases. Semiparametric regression models are also introduced and used to summarize treatment effects; we give weighted estimation procedures that adapt for event-dependent censoring. Some concluding remarks are given in Section 7.

2 Notation and Specification of Treatment Effects

Let $\{N(t), 0 < t\}$ denote the right continuous process counting the number of events over the time interval $(0, t]$ for a general individual, and let $dN(t) = 1$ if an event occurs at time t and $dN(t) = 0$ otherwise. The time of the kth event is denoted by T_k, $k = 1, 2, \ldots$. The history of the process contains information on the number and times of prior events at a particular time and is denoted by $H(t) = \{N(u), 0 < u < t\}$. In practice, processes are observed over a finite period of time, so we also let C be a right censoring or end-of-follow-up time and $C(t) = I(t \leq C)$ indicate whether the process is under observation at time t.

Recurrent event processes can be specified through the event intensity function (Cook and Lawless, 2007, Sec. 1.3)

$$\lambda(t|H(t)) = \lim_{\Delta t \to 0} \frac{P\left(\Delta N(t) = 1 | H(t)\right)}{\Delta t},$$

where $\Delta N(t) = N(t + \Delta t^-) - N(t^-)$ is the number of events occurring over the interval $[t, t + \Delta t)$. For continuous time processes where at most one event can occur at any time, this function fully defines the process. Given the intensity, the likelihood function based on observed sets of event times across independent individuals is readily written down (Cook and Lawless, 2007, Sec. 2.6), and the validity of maximum likelihood estimation is maintained even when censoring is related to event history.

If the probability of a new event at time t depends substantially on the previous event history, then specifying the effects of treatments in terms of the intensity is not especially desirable, as argued in Section 1. In particular, a treatment that affects the event history $H(t)$ will be confounded with $H(t)$ as far as the intensity $\lambda(t|H(t))$ is concerned, thus rendering interpretation difficult. Moreover, it is desirable to define treatment effects in terms of outcomes that are easily interpreted at the start of the study, or at randomization. As argued in Section 1, this suggests that treatment effects be specified in terms of marginal features such as the mean function $\mu(t)$. For example, the ratio or difference of mean functions for two treatment groups is easily interpreted.

When interest lies in marginal features, alternative approaches are, in principle, available for analysis. If intensity functions giving a full specification for the event process are specified for different treatment groups, then maximum likelihood methods can be applied and marginal features can then be obtained based on this full model. However, these are typically complicated functions of model parameters and treatment covariates. A simpler and widely adopted alternative is to estimate marginal features directly via simple empirical estimators, giving easily interpreted measures of treatment effect. However, this typically involves no conditioning on the process history and such estimates can be inconsistent under event-dependent censoring. We discuss this in Section 3 and show how inverse probability of censoring weights can be used in this case to obtain consistent estimates of marginal features under correct specification of a model for the censoring process. We also consider various methods of estimating marginal mean or rate functions, which differ in the extent to which they condition on event history, and in the need for inverse probability of censoring weights.

3 ESTIMATION OF MEAN FUNCTIONS AND OTHER MARGINAL FEATURES

3.1 METHODS BASED ON MARGINAL RATE FUNCTIONS

We deal in this section with the case where there is no terminating event. Let $\{N_i(t), 0 < t\}$ be the counting process for subject i, C_i the censoring time, and $C_i(t) = I(t \leq C_i)$ the indicator function of whether individual i is under observation at t. We consider both discrete and continuous time processes and let $\mu(t)$ be the nondecreasing, right-continuous mean function. If $\mu(t)$ is continuous at t, we write $d\mu(t) = \rho(t)dt$; if $\mu(t)$ jumps at t, then $d\mu(t)$ is the jump size.

We consider nonparametric estimation of $\mu(t)$, although a similar development applies to parametric models. Covariates are not considered here, as we assume initially that separate estimates will be obtained for different treatment groups. As do many other authors (e.g., Lin et al., 2000; Miloslavsky et al., 2004) we write, informally, $E\{dN_i(t)\} = d\mu(t)$ and, correspondingly, $E\{dN_i(t) - d\mu(t)\} = 0$ for both the discrete and continuous time cases. If observations are available on independent individuals $i = 1, 2, \ldots, m$, we then consider the estimating equations

$$\sum_{i=1}^{m} C_i(t)\{dN_i(t) - d\mu(t)\} = 0\,, \tag{3.1}$$

for different values of t. Because our nonparametric estimates of $\mu(t)$ are discrete with jumps only at event times, there is no harm in doing this. The Equation (3.1) is formally justified under a Poisson model as maximum likelihood equations (Lawless and Nadeau, 1995), but are unbiased more generally and may be used more generally to estimate the mean function. They give

$$d\widehat{\mu}(t) = \frac{\sum_{i=1}^{m} C_i(t)dN_i(t)}{\sum_{i=1}^{m} C_i(t)} \tag{3.2}$$

and the resulting estimate of the mean function $\widehat{\mu}(t) = \int_0^t d\widehat{\mu}(s)$ is the Nelson-Aalen estimate (e.g., Andersen et al., 1993, Sec. 4.1). Note that, unlike the case of survival data, $C_i(t)$ is known for all individuals at all times $t > 0$.

Lawless and Nadeau (1995) justify (3.1) more formally and provide discrete-time asymptotics. Lin et al. (2000) give a rigorous treatment of the continuous time case. The validity of (3.1) and $\widehat{\mu}(t)$ for general processes requires that C_i be completely independent of $\{N_i(t), 0 < t\}$ in which case $E\{dN_i(t)|C_i \geq t\} = E\{dN_i(t)\} = d\mu(t)$. This is a much stronger condition than conditionally independent censoring necessary for intensity-based analysis, which requires only $E\{dN_i(t)|C_i \geq t, H_i(t)\} = E\{dN_i(t)|H_i(t)\}$.

The estimating Equation (3.1) can be modified by the introduction of inverse probability of censoring weights originated by Robins and Rotnitzky (1992) and Robins (1993) in the context of survival analysis. Van der Laan et al. (2002) and Miloslavsky et al. (2004) give recent discussions for general event history processes. We proceed as follows. Assume that the conditional hazard function for the censoring at time t, given the entire event history $H_i(\infty)$ for an individual, depends only on the event history $H_i(t)$ up to time t (e.g., see Satten et al., 2001; van der Laan et al., 2002). Then let $G_i(t) = P(C_i \geq t|H_i(\infty)) = P(C_i \geq t|H_i(t))$ and consider the estimating equation

$$\sum_{i=1}^{m} U_i(t) = \sum_{i=1}^{m} \frac{C_i(t)}{G_i(t)}\{dN_i(t) - d\mu(t)\} = 0\,. \tag{3.3}$$

Then $E\{U_i(t)\} = 0$ in (3.3) because $E_{H_i(t)}\{E[U_i(t)|H_i(t)]\} = E_{H_i(t)}\{dN_i(t) - d\mu(t)\} = 0$. This development also extends to cases where censoring depends on external covariates, under the assumption that the censoring intensity at t depends only on covariate values up to that time (e.g., Miloslavsky et al., 2004).

To use (3.3), $G_i(t)$ must be replaced with an estimate $\widehat{G}_i(t)$. Let $\lambda^c(t|H_i(t))$ and $\Lambda^c(t|H_i(t)) = \int_0^t \lambda^c(u|H_i(u))du$ denote the censoring intensity (conditional hazard) function and cumulative intensity functions respectively. Then note that by product integration (Andersen et al., 1993, Sec. 2.6)

$$G_i(t) = \prod_{u<t} \{1 - d\Lambda^c(u|H_i(u))\} \tag{3.4}$$

where $d\Lambda^c(t|H_i(t)) = \lambda^c(t|H_i(t))dt$ if the intensity is continuous at t. One type of dependent censoring we consider here takes the censoring intensity as a function of t and $N_i(t^-)$, with

$$d\Lambda^c(t|H_i(t)) = d\Lambda^c(t|N_i(t^-)) = d\Lambda^c_j(t) \tag{3.5}$$

if $N_i(t^-) = j$. In this case the intensity for censoring depends only on the cumulative number of events experienced by t^-; this is sometimes called event-dependent, or more generally state-dependent (Datta and Satten, 2002; Gunnes et al., 2007) censoring. A nonparametric estimate of (3.5) is

$$d\widehat{\Lambda}^c_j(t) = \frac{\sum_{i=1}^m I(C_i = t, N_i(t^-) = j)}{\sum_{i=1}^m I(C_i \geq t, N_i(t^-) = j)} \tag{3.6}$$

and then $d\widehat{\Lambda}^c(t|H_i(t)) = \sum_{j=0}^\infty d\widehat{\Lambda}^c_j(t) I(N_i(t^-) = j)$, which can be inserted into (3.4) to get $\widehat{G}_i(t)$. Equation (3.3) is then modified by inserting this estimate in place of $G_i(t)$, giving the weighted rate function estimate

$$d\widehat{\mu}(t) = \frac{\sum_{i=1}^m C_i(t) dN_i(t)/\widehat{G}_i(t)}{\sum_{i=1}^m C_i(t)/\widehat{G}_i(t)} \tag{3.7}$$

and the weighted Nelson-Aalen estimate $\widehat{\mu}(t) = \int_0^t d\widehat{\mu}(u)$. Variance estimation for unweighted $\widehat{\mu}(t)$ is discussed by Lawless and Nadeau (1995) and Lin et al. (2000). For the IPCW case it may be possible to derive variance estimates, but these would be very complicated and it is simpler to employ bootstrap methods.

3.2 Estimation via Marginal Failure Time Models

Another approach to estimation of the mean function is to note that

$$\mu(t) = \sum_{k=1}^\infty k P_{0k}(0,t) \tag{3.8}$$

where $P_{0k}(t) = P(N(t) = k)$. Thus we first estimate the probabilities $P_{0k}(t)$ of k events over $(0,t]$, which are another marginal feature of interest. One way this can be done is along the lines discussed by Pepe (1991), Pepe et al. (1991), and Couper and Pepe (1997) in slightly different contexts. In particular, we note that

$$\begin{aligned} P_{0k}(0,t) &= P(T_k \leq t) - P(T_{k+1} \leq t) \\ &= P(T_{k+1} > t) - P(T_k > t)\,, \end{aligned}$$

where T_k is the time of the kth event. The idea is then to estimate $P(T_k > t)$ by a standard Kaplan-Meier estimate. Thus let T_{ki} be the time of the kth event and $N_{ki}(t) = I(T_{ki} \leq t)$ be the counting process for the kth event for individual i, $i = 1, \ldots, m$. Let $Y_{ki}(s) = I(s \leq T_{ki})$ indicate that individual i has not yet experienced their kth event at time s^- and $\bar{Y}_{ki}(s) = C_i(s)Y_{ki}(s)$ indicate that individual i is under observation and has not yet experienced their kth event.

Let $H_k(t)$ denote the cumulative hazard function for T_k, and let $\mathcal{F}_k(t) = P(T_k > t)$. Then the estimating functions

$$\sum_{i=1}^{m} \bar{Y}_{ki}(t)\{dN_{ki}(t) - dH_k(t)\} = 0 \tag{3.9}$$

give the estimates

$$d\widehat{H}_k(t) = \frac{\sum_{i=1}^{m} \bar{Y}_{ki}(t)dN_{ki}(t)}{\sum_{i=1}^{m} \bar{Y}_{ki}(t)} \tag{3.10}$$

which can be used to obtain the Kaplan-Meier estimates

$$\widehat{\mathcal{F}}_k(t) = \prod_{(0,t]} \{I - d\widehat{H}_k(u)\}$$

by product integration. A consistent estimate for $\mu(t)$ is obtained by substituting $\widehat{P}_{0k}(t) = \widehat{\mathcal{F}}_{k+1}(t) - \widehat{\mathcal{F}}_k(t)$ in (3.8), which we call the Pepe estimate.

Equation (3.9) is, like (3.1), only valid under independent censoring, and an IPCW version of (3.9) is

$$\sum_{i=1}^{m} \frac{C_i(t)}{\widehat{G}_i(t)} Y_{ki}(t)\{dN_{ki}(t) - dH_k(t)\} = 0 \tag{3.11}$$

where $\widehat{G}_i(t)$ is estimated as in Section 3.1. This gives

$$d\widehat{H}_k(t) = \frac{\sum_{i=1}^{m} \bar{Y}_{ki}(t)dN_{ki}(t)/\widehat{G}_i(t)}{\sum_{i=1}^{m} \bar{Y}_{ki}(t)/\widehat{G}_i(t)}\,, \tag{3.12}$$

and a weighted marginal survivor function of the form

$$\widehat{\mathcal{F}}_k(t) = \prod_{(0,t]} \{1 - d\widehat{H}_k(u)\}\,, \tag{3.13}$$

as described in Satten et al. (2001). Using this estimate of the survivor function gives a weighted Pepe estimate. In practice, one can employ bootstrap sampling for variance estimation.

3.3 Estimates Based on Multistate Models

An alternative approach is to adopt a Markov multistate model as a basis for nonparametric estimation. We consider a model with states $E_0, E_1, E_2, \ldots$, where an individual is in state E_j at time t if $N_i(t) = j$, and let $V_{ki}(t) = I(N_i(t^-) = k - 1)$ be an "at risk" indicator for transitions into state k. In contrast to the indicator $Y_{ki}(t)$ discussed in Section 3.2, $V_{ki}(t)$ indicates whether individual i is truly at risk of their kth event by requiring them to be in state $k - 1$; we also define $\bar{V}_{ki}(t) = C_i(t)V_{ki}(t)$. If the

multistate process is Markov, then we may denote the intensity for a transition from state $k-1$ to state k at time $t>0$ as $\alpha_k(t)$ for $k=1,2,\ldots$, and let $A_k(t)=\int_0^t \alpha_k(u)du$.

A nonparametric estimate of the transition probability matrix $\boldsymbol{P}(s,t)$, with entries $P_{\ell k}(s,t) = P(N_i(t)=k|N_i(s)=\ell)$, for $0\le s\le t$, is given by the Aalen-Johansen estimate (Andersen et al., 1993, Sec. 4.4). For clarity of exposition, we briefly outline it here, as well as the refinements made to it by Datta and Satten (2002). Let $\boldsymbol{A}(t)$ be the upper triangular matrix with entries $-A_j(t)$, $j=1,2,\ldots$, on the diagonal, entries $A_j(t)$ in the superdiagonal positions (i.e., $\boldsymbol{A}_{j,j+1}(t)$), and zero elsewhere. The matrix of transition probabilities is then given by (see Andersen et al., 1993, Sec 2.6) the product integral

$$\boldsymbol{P}(s,t)=\prod_{(s,t]}\{I+d\boldsymbol{A}(u)\}\ . \tag{3.14}$$

The nonparametric estimates

$$d\widehat{A}_k(t)=\frac{\sum_{i=1}^m \bar{V}_{ki}(t)dN_{ki}(t)}{\sum_{i=1}^m \bar{V}_{ki}(t)} \tag{3.15}$$

yield the Aalen-Johansen estimate

$$\widehat{\boldsymbol{P}}(s,t)=\prod_{(s,t]}\left\{I+d\widehat{\boldsymbol{A}}(u)\right\} \tag{3.16}$$

(Andersen et al., 1993, Sec. 4.4). When the process is Markov, it is a consistent estimator of $\boldsymbol{P}(s,t)$ when the number of states is finite; we can achieve this in the present case by imposing a limit K on the number of recurrent events deemed possible over the follow-up period in question. We remark that in finite samples there may be cases where the denominator in (3.15) equals zero. In that case the ratio $0/0$ in (3.15) is defined to be zero; this yields appropriate estimates (3.16) for transition probabilities. A similar convention is used for estimates in Section 4.

Aalen et al. (2001) and Datta and Satten (2001) have pointed out that the Aalen-Johansen estimator is also consistent for state occupancy probabilities $\boldsymbol{P}(0,t)$ for non-Markov multistate processes under independent censoring. This implies that the corresponding "Aalen-Johansen" (AJ) estimate of the mean function obtained using (3.8) with the corresponding entry of the AJ estimate $\widehat{\boldsymbol{P}}(0,t)$ used in place of $P_{0k}(t)$, is also robust.

The AJ estimate of $\mu(t)$ is thus robust if censoring times C_i are independent of the event process or, more generally, if the recurrent event process is Markov. However, Datta and Satten (2002) have shown that the AJ estimate of $\boldsymbol{P}(s,t)$ can be adjusted by the IPCW method, so as to make it robust when there is event-dependent censoring.

Defining the $\widehat{G}_i(t)$ as in Section 3.1, Datta and Satten (2002) give IPCW estimates

$$d\widehat{A}_k(t)=\frac{\sum_{i=1}^n \bar{V}_{ki}(t)dN_{ki}(t)/\widehat{G}_i(t)}{\sum_{i=1}^n \bar{V}_{ki}(t)/\widehat{G}_i(t)} \tag{3.17}$$

and use them in (3.16) to provide an IPCW estimate of the transition probability matrix. Provided the censoring model used to obtain the $G_i(t)$ is correct, this gives a consistent estimator of $\boldsymbol{P}(0,t)$. A consistent estimate of $\mu(t)$ is then obtained by using this in (3.8). We will refer to this estimate in what follows as the Datta-Satten or weighted Aalen-Johansen estimate of $\mu(t)$. Datta and Satten (2002) give variance estimates for $\widehat{P}(0,t)$ but they are very awkward to apply to $\widehat{\mu}(t)$ and so we suggest using bootstrap resampling methods.

4 MARGINAL FEATURES WITH TERMINATING EVENTS

4.1 ESTIMATES BASED ON MARGINAL RATE AND SURVIVAL FUNCTIONS

Recurrent events frequently arise in settings where individuals are at high risk of a terminating event, the setting motivating this article being a prime example. Cook and Lawless (1997) discuss marginal features in this setting. For convenience, we refer to the terminating event here as death, and let T_{ki} denote the time of the kth recurrent event as before and let T_i^d denote the time of death for individual i, $i = 1, \ldots, m$. If $d\mu^*(u) = E\{dN_i(u)|T_i^d \geq u\}$ denotes the rate of events at u conditional on being alive at time u^-, and $\mathcal{F}^d(u) = P(T_i^d \geq u)$ is the survivor function for T_i^d, then

$$E\{N_i(t)\} = \int_0^t \mathcal{F}^d(u)\, d\mu^*(u) \tag{4.1}$$

is the marginal expected number of events accounting for the possible realization of the terminating event which precludes subsequent recurrent events.

The counting processes $\{N_i(t), 0 < t\}$ and $\{N_i^d(t), 0 < t\}$ record the recurrent events and death respectively so $dN_i(t) = 1$ if an event occurs at time t for individual i and $dN_i(t) = 0$ otherwise, and $dN_i^d(t) = 1$ if death occurs at time t for individual i and $dN_i^d(t) = 0$ otherwise. Again C_i is a right censoring time, $C_i(t) = I(t \leq C_i)$, and $Y_i^d(t) = I(t \leq T_i^d)$ indicates whether individual i is alive at t.

An estimate of (4.1) can be obtained by modifying (3.3) to

$$\sum_{i=1}^m \frac{C_i(t)}{G_i(t)} Y_i^d(t)\{dN_i(t) - d\mu^*(t)\} = 0\,,$$

leading to

$$d\widehat{\mu}^*(t) = \frac{\sum_{i=1}^m C_i(t)Y_i^d(t)dN_i(t)/\widehat{G}_i(t)}{\sum_{i=1}^m C_i(t)Y_i^d(t)/\widehat{G}_i(t)}\,. \tag{4.2}$$

Also note that a weighted Kaplan-Meier estimate for the survivor function that allows for event-dependent censoring is obtained by solving

$$\sum_{i=1}^m \frac{C_i(t)}{\widehat{G}_i(t)} Y_i^d(t)\{dN_i^d(t) - dH^d(t)\} = 0\,, \tag{4.3}$$

to get

$$d\widehat{H}^d(t) = \sum_{i=1}^m \frac{C_i(t)Y_i^d(t)dN_i^d(t)}{\widehat{G}_i(t)} \Big/ \sum_{i=1}^m \frac{C_i(t)Y_i^d(t)}{\widehat{G}_i(t)}\,, \tag{4.4}$$

and

$$\widehat{\mathcal{F}}^d(t) = \prod_{(0,t]} \{I - d\widehat{H}^d(u)\} \tag{4.5}$$

(Satten et al., 2001). Using the estimates (4.2) and (4.5) in (4.1) gives a weighted version of the estimator of $\mu(t)$ proposed by Cook and Lawless (1997); using $\widehat{G}_i(t) = 1$ gives their actual estimator. We refer to this estimator of $\mu(t)$ in what follows as the weighted Cook-Lawless estimator.

4.2 Estimation via Cumulative Incidence Functions

The analogous approach to the Pepe method of Section 3.2 requires estimation of cumulative incidence functions because death is a competing risk for occurrence of the kth event (Crowder, 2001; Kalbfleisch and Prentice, 2002). As before $C_i(s) = I(s \leq C_i)$ and $Y_i^d(s) = I(s \leq T_i^d)$, and $Y_{ki}(s) = I(s \leq T_{ki})$ indicates that individual i has not experienced the kth event. We define $dM_{ki}^d(s) = 1$ if individual i dies at time s without having experienced their kth event and define it as being zero otherwise, and let $Y_{ki}^d(s) = 1 - M_{ki}^d(s^-)$; the corresponding counting process for $dM_{ki}^d(t)$ is $\{M_{ki}^d(s), 0 < s\}$.

The cumulative incidence function for the kth event in the continuous time case is

$$F_k(t) = P(T_{ki} \leq t) = \int_0^t dH_k(u) \exp(-(H_k(u) - H_k^d(u)) \tag{4.6}$$

where $H_k(u)$ and $H_k^d(u)$ are respectively, in competing risks terminology, the cumulative cause-specific hazards of the kth event and death before the kth event. The associated weighted estimating equations are

$$\sum_{i=1}^m \frac{C_i(t)}{\widehat{G}_i(t)} Y_{ki}(t) Y_{ki}^d(t) \{dN_{ki}(t) - dH_k(t)\} = 0 \tag{4.7}$$

and

$$\sum_{i=1}^m \frac{C_i(t)}{\widehat{G}_i(t)} Y_{ki}(t) Y_{ki}^d(t) \{dM_{ki}^d(t) - dH_k^d(t)\} = 0 \tag{4.8}$$

for $dH_k(t)$ and $dH_k^d(t)$ respectively. We then plug the resulting discrete estimates into (4.6) to give $\widehat{F}_k(t)$ (Lawless, 2002, Sec. 9.2).

Consider the multistate process diagram in Figure 2 and let $Z(t)$ be the state occupied by an individual at time t. Transitions to the right (i.e., $E_k \rightarrow E_{k+1}$) correspond to the occurrence of new events and downward transitions (i.e., $E_k \rightarrow D_k$) correspond to deaths; because the death states are absorbing, the latter transitions determine the lifetime number of events. The probability of experiencing exactly k events over $(0, t]$ is then $P(Z(t) = E_k \text{ or } D_k)$ which is simply $F_k(t) - F_{k+1}(t)$. An estimate of $\mu(t)$ based on this formulation is therefore

$$\widehat{\mu}(t) = \sum_{k=0}^{\infty} k[\widehat{F}_k(t) - \widehat{F}_{k+1}(t)] \, . \tag{4.9}$$

We refer to this estimate as the weighted ($G_i(t)$ estimated) and unweighted ($G_i(t) = 1$) Pepe estimate in the setting with a terminal event.

4.3 Estimates Based on Multistate Models

We continue to work with the multistate process of Figure 2. Here we define $\{N_{ki}(t), 0 < t\}$ as the counting process for kth events as before but also define $\{N_{ki}^d(t), 0 < t\}$ as the counting process recording the number of $E_k \rightarrow D_k$ transitions (i.e., deaths from state E_k). As in Section 3.3 we let $V_{ki}(t) = I(Z(t^-) = E_{k-1})$ indicate whether individual i is at risk of a kth event, and let $\bar{V}_{ki}(s) = C_i(s) V_{ki}(s)$. Finally, we make the working assumption that the process is Markov, and define $A_k(t) = \int_0^t \alpha_k(u) du$ and $A_k^d(t) = \int_0^t \alpha_k^d(u) du$ as the cumulative intensity functions for transitions from E_{k-1} to E_k and E_k to D_k, respectively.

In this case we specify a (conveniently large) upper limit K on the number of recurrent events that can occur, so that $\alpha_{K+1}(t) = 0$. The IPCW estimates for $dA_k(t)$ and $dA_k^d(t)$ are

$$d\widehat{A}_k(t) = \frac{\sum_{i=1}^{m} \bar{V}_{ki}(t) dN_{ki}(t)/\widehat{G}_i(t)}{\sum_{i=1}^{m} \bar{V}_{ki}(t)/\widehat{G}_i(t)} \, , k = 1, 2, \ldots, K \tag{4.10}$$

and

$$d\widehat{A}_k^d(t) = \frac{\sum_{i=1}^{m} \bar{V}_{k+1,i}(t) dN_{ki}^d(t)/\widehat{G}_i(t)}{\sum_{i=1}^{m} \bar{V}_{k+1,i}(t)/\widehat{G}_i(t)} \, , k = 0, 1, 2, \ldots, K \tag{4.11}$$

respectively, where $\widehat{G}_i(t)$ is defined as in previous sections.

Writing the $2K + 2$ states in order $E_0, E_1, \ldots, E_K$, $D_0, D_1, \ldots, D_K$, we define the $(2K + 2) \times (2K+2)$ matrix $d\widehat{\boldsymbol{A}}(t)$ with entries $d\widehat{A}_{E_k E_k}(u) = -d\widehat{A}_k(u) - d\widehat{A}_k^d(u)$, $k = 0, 1, \ldots, K$, $d\widehat{A}_{E_k E_{k+1}}(u) = d\widehat{A}_k(u)$, $k = 0, 1, \ldots, K-1$, and $d\widehat{A}_{E_k D_k}(u) = d\widehat{A}_k^d(u)$, $k = 0, 1, \ldots, K$. The Datta-Satten (or weighted Aalen-Johansen) estimate of $\boldsymbol{P}(0, t)$ is then given by

$$\widehat{\boldsymbol{P}}(0, t) = \prod_{(0,t]} \left\{ I + d\widehat{\boldsymbol{A}}(u) \right\} \, .$$

Defining $P_{E_0 E_k}(0, t) = P(Z(t) = E_k | Z(0) = E_0)$, $P_{E_0 D_k}(0, t) = P(Z(t) = D_k | Z(0) = E_0)$, and $P_{0k}(t) = P_{E_0 E_k}(0, t) + P_{E_0 D_k}(0, t)$, then

$$\widehat{\mu}(t) = \sum_{k=1} k \widehat{P}_{0k}(t) \tag{4.12}$$

is a Datta-Satten, or weighted Aalen-Johansen, estimate of the mean function; using $\widehat{G}_i(t) = 1$ gives an unweighted Aalen-Johansen estimate in this setting with terminal events.

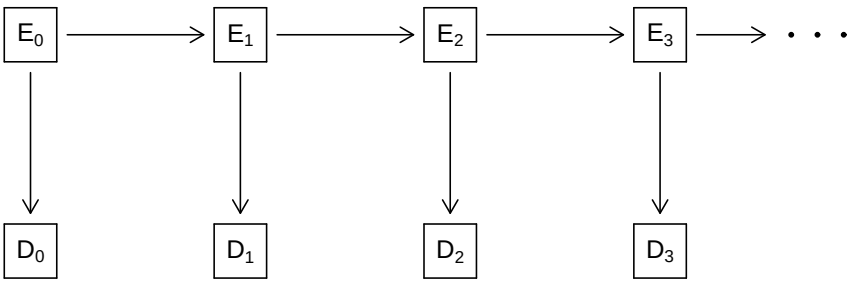

Figure 2: Multistate diagram for recurrent events and death.

5 NUMERICAL STUDIES

5.1 SIMULATIONS FOR RECURRENT EVENTS

The multistate diagram in Figure 2 reflects the possible course of events for an individual, where E_k is the state occupied for subjects alive who have experienced k events, and D_k is the state occuped for subjects having experienced k events before death. In general, let us denote the recurrent event intensity as $\lambda(t|H(t))$ and let $\lambda^d(t|H(t))$ be the intensity for death. The history $H(t)$ at time t determines what state in Figure 2 an individual occupies at time t. We begin by considering settings without a terminal event as in Section 3 so that $\lambda^d(t|H(t)) = 0$.

Simulation studies were conducted to investigate the frequency properties of unweighted and weighted Nelson-Aalen, Pepe, and Aalen-Johansen estimators for Markov and and non-Markov (semi-Markov) processes under both independent and state-dependent censoring. For the Markov processes, we let $\lambda(t|H(t)) = \lambda_0 E^k$, if $N_i(t^-) = k$, $k = 0, ..., 4$ where λ_0 is the intensity for the first event, and E is a multiplicative factor that reflects how the risk of the first few subsequent events changes with each additional event; we set $\lambda(t|H(t)) = \lambda_0 E^4$ for $N(t^-) \geq 5$. To examine the robustness of the estimators to departures from the Markov assumption we also generated data from a semi-Markov process in which $\lambda(t|H(t)) = \zeta_k(B(t))$ where $B(t)$ is the time since the last event (or "backwards recurrence time") and $N(t^-) = k$, and $\zeta_k(\cdot)$ is the hazard function from a gamma distribution with density function $f(w) = w^{\alpha-1}\beta_k^\alpha \exp(-\beta_k w)/\Gamma(\alpha)$. We set the parameters of the gamma distribution as shape $\alpha = 2$ and inverse scale $\beta_k = E^k\beta$, $k = 0, \ldots, 4$ and $\beta_k = E^4\beta$ for $k > 4$. Thus, as in the Markov process, the event intensity changed for the first four events and remained constant thereafter.

The preceding processes were observed subject to administrative censoring at $\tau = 1$, as well as to earlier, possibly state-dependent, censoring. We refer to the units of t as years for convenience. We considered a Markov model for censoring, and let $\lambda^c(t|H(t)) = \lambda_0^c C^k$ denote the rate of censoring from state k, $k = 0, 1, \ldots$, where λ_0^c is the censoring rate in state E_0 and C is a multiplicative factor that indicates how the rate of censoring changes with the cumulative number of events.

The parameter values for the Markov process were determined by fixing E and solving for λ_0 so that $E\{N(1)\}$ was 1 or 2 and for the semi-Markov process β was determined to give these same expectations. To obtain the desired censoring rates, the value of C was then specified, and using the explicit solution via the Chapman-Kolmogorov equations (Cox and Miller, 1965), λ_0^c was determined to give probabilities of censoring before $\tau = 1$ of 25% or 50%. We set the multiplicative factors to be $E = 1$ and 1.5 for the Markov model, and set $C = 1$ for independent censoring and $C = 1.5$ for state-dependent censoring for both models. For the semi-Markov model we set $E = 1$ and considered $C = 1$ and $C = 1.5$.

One thousand datasets of $m = 500$ subjects were simulated for each model and parameter configuration. This number of subjects is realistic in many randomized trials involving recurrent event responses, and 1000 simulations is enough to provide a clear picture of the bias and variability of the different estimators. The mean function $\mu(t)$ was estimated using the unweighted and weighted Nelson-Aalen estimate, the unweighted and weighted Pepe estimate, and the unweighted and weighted Aalen-Johansen (Datta-Satten) estimate. The estimate of $G_i(t)$ was based on the most general event-dependent censoring model with censoring rates estimated as in (3.6). Figure 3 gives plots of the empirical bias over $(0, 1]$ for the six estimates for the Markov and semi-Markov models respectively with $\mu(1) = 2$, 50% censoring, $C = 1.5$, and $E = 1.5$. With 1000 simulations, the standard deviations are no more than about 0.004 for the biases shown.

From Figure 3 we can see that under the Markov process, the unweighted Pepe and Nelson-Aalen estimates of $\mu(t)$ featured the greatest bias with both methods giving increasingly conservative estimates of the expected number of events over time. The unweighted Aalen-Johansen estimate performed much better than the other unweighted estimators, as one would expect because the Markov model adapts to this form of state-dependent censoring. The weighted Pepe and Nelson-Aalen meth-

ods performed considerably better with negligible bias even at $\tau = 1$; the plot of the weighted Aalen-Johansen estimate is superimposed on the plot of the unweighted estimate. For the semi-Markov process, all unweighted methods had some bias with the Nelson-Aalen estimate overestimating the mean slightly and the Aalen-Johansen method underestimating it; weighted versions for all three methods performed very well.

The empirical biases and standard errors of the estimators at $\tau = 1$ with $C = 1$ and $C = 1.5$ are given in Tables 2 and 3 for the Markov and semi-Markov settings respectively. Under independent censoring ($C = 1$) the empirical standard errors are remarkably similar across the configurations examined. For Markov processes subject to state-dependent censoring, the weighted Pepe estimators had very slightly higher variability compared with the weighted Nelson-Aalen and weighted Aalen-Johansen estimators when the mean number of events was two. The smaller empirical biases and low empirical variabilities suggest the weighted Nelson-Aalen and weighted Aalen-Johansen approaches may be preferred. Similar findings were seen for the semi-Markov setting in Table 3; the weighted estimators typically had negligible bias and reasonably similar efficiency.

We also simulated data under a semi-Markov censoring model in which $\lambda^c(t|H(t)) = [\lambda_0^c C^k]^{I(B(t)<\Delta)}$, $k = 0, \ldots, 4$ which means that the censoring intensity is zero other than at times within Δ of an event, and the value for the censoring intensity when it is nonzero changes with the occurence of each event up the fourth event; $\lambda^c(t|H(t)) = [\lambda_0^c C^4]^{I(B(t)<\Delta)}$ for $k = 5, \ldots$. To examine the impact of misspecification of the censoring process the IPCW estimates were used with the Markov censoring model of (4.5). The results, displayed in Table 4, indicate that the bias in the unweighted estimators was quite substantial for the Nelson-Aalen and Pepe approaches, but tended to be smaller for the Aalen-Johansen estimator. Even though the model for the censoring process is misspecified, the weighted methods give much lower bias for the Nelson-Aalen and Pepe approaches; the empirical standard errors were at times much smaller for the weighted Nelson-Aalen approach than the Pepe approach.

Table 2: Empirical biases and standard errors for unweighted and weighted estimators under a Markov model ($\mu = E\{N(1)\}$, $m = 500$, 1000 simulations)

				C = 1.0						C = 1.5					
μ	E	% CENS		NA	WNA	PEPE	WPEPE	AJ	WAJ	NA	WNA	PEPE	WPEPE	AJ	WAJ
1	1	25	BIAS	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	-0.010	0.002	0.003	0.003
			ESE	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.049)	(0.049)	(0.048)	(0.049)	(0.049)	(0.049)
		50	BIAS	0.004	0.004	0.004	0.003	0.004	0.003	0.004	0.004	-0.026	0.002	0.003	0.003
			ESE	(0.057)	(0.057)	(0.058)	(0.057)	(0.057)	(0.057)	(0.055)	(0.055)	(0.052)	(0.055)	(0.055)	(0.055)
1	1.5	25	BIAS	0.003	0.002	0.003	0.002	0.003	0.003	-0.016	-0.000	-0.029	-0.001	0.001	0.001
			ESE	(0.066)	(0.066)	(0.067)	(0.066)	(0.066)	(0.066)	(0.062)	(0.065)	(0.061)	(0.065)	(0.065)	(0.065)
		50	BIAS	0.005	0.003	0.006	0.003	0.004	0.004	-0.036	-0.004	-0.063	-0.007	-0.001	-0.002
			ESE	(0.074)	(0.074)	(0.075)	(0.074)	(0.074)	(0.074)	(0.068)	(0.075)	(0.065)	(0.076)	(0.075)	(0.076)
2	1	25	BIAS	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	-0.036	0.000	0.001	0.001
			ESE	(0.068)	(0.068)	(0.069)	(0.068)	(0.068)	(0.068)	(0.066)	(0.066)	(0.065)	(0.066)	(0.066)	(0.066)
		50	BIAS	-0.001	-0.001	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.087	-0.005	-0.002	-0.003
			ESE	(0.077)	(0.077)	(0.079)	(0.077)	(0.077)	(0.077)	(0.076)	(0.077)	(0.072)	(0.077)	(0.077)	(0.077)
2	1.5	25	BIAS	0.003	0.002	0.003	0.002	0.003	0.003	-0.067	-0.007	-0.107	-0.012	-0.003	-0.004
			ESE	(0.112)	(0.111)	(0.113)	(0.111)	(0.111)	(0.111)	(0.099)	(0.107)	(0.096)	(0.108)	(0.107)	(0.108)
		50	BIAS	0.003	-0.001	0.005	-0.002	0.001	0.001	-0.148	-0.029	-0.230	-0.041	-0.023	-0.023
			ESE	(0.118)	(0.116)	(0.121)	(0.116)	(0.116)	(0.116)	(0.103)	(0.120)	(0.097)	(0.128)	(0.120)	(0.120)

5.2 Simulations for Recurrent and Terminal Events

Problems with terminal events were simulated by setting $\lambda^d(t|H(t)) = \lambda_0^d D^k$ with $\lambda_0^d > 0$. Under a Markov process we specified the form of the event intensities as in Section 5.1. The values for

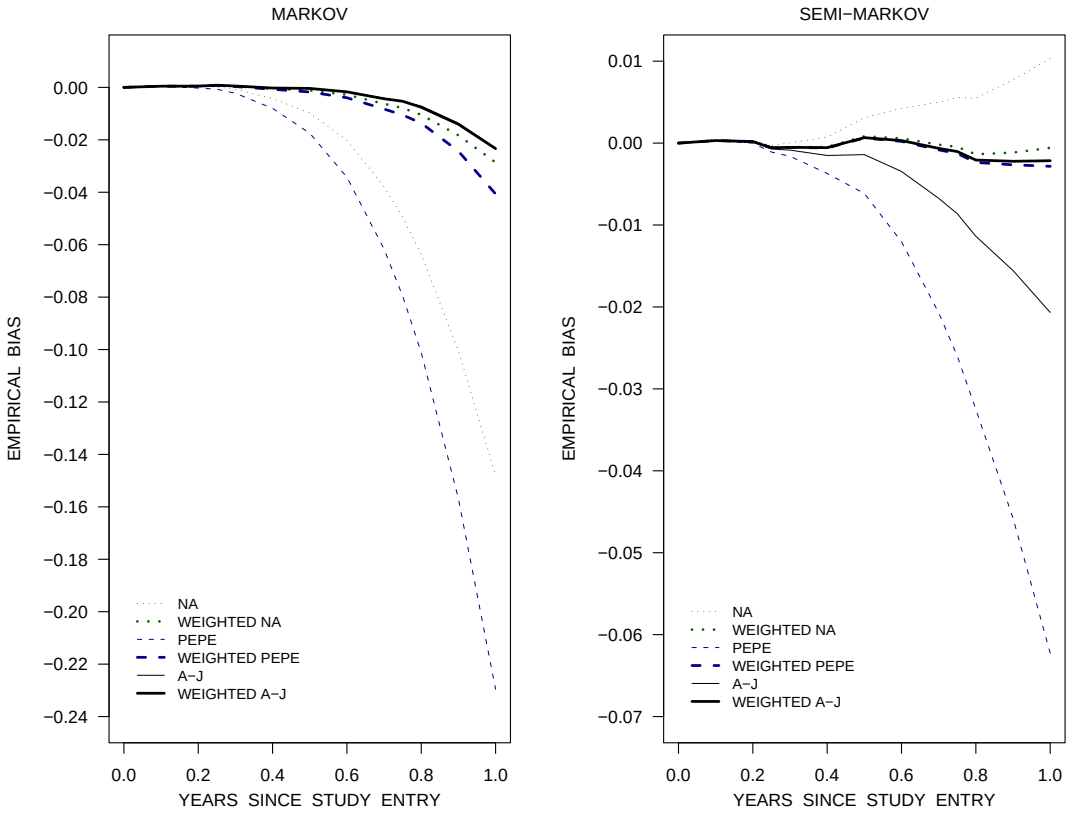

Figure 3: Empirical biases of unweighted and weighted estimates of mean function over time for Markov and semi-Markov processes without termination (50% censoring at $\tau = 1$ with $C = 1.5$, $E = 1.5$ for Markov model; $\alpha = 2$ for semi-Markov model, $\mu(1) = 2$).

Table 3: Empirical biases and standard errors for unweighted and weighted estimators under a semi-Markov model ($\alpha = 2$, $m = 500$, 1000 simulations)

				C = 1.0						C = 1.5					
μ	E	% CENS		NA	WNA	PEPE	WPEPE	AJ	WAJ	NA	WNA	PEPE	WPEPE	AJ	WAJ
1	1.0	25	BIAS	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.003	-0.000	-0.010	-0.001	-0.003	-0.001
			ESE	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
		50	BIAS	-0.001	-0.001	-0.002	-0.001	-0.001	-0.001	0.007	-0.000	-0.022	-0.001	-0.007	-0.001
			ESE	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.044)	(0.046)	(0.045)	(0.046)
1	1.5	25	BIAS	0.000	-0.000	0.000	-0.000	-0.000	-0.000	-0.011	-0.002	-0.023	-0.003	-0.007	-0.002
			ESE	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.053)	(0.054)	(0.052)	(0.054)	(0.054)	(0.055)
		50	BIAS	-0.002	-0.003	-0.002	-0.003	-0.003	-0.003	-0.026	-0.006	-0.051	-0.008	-0.017	-0.006
			ESE	(0.062)	(0.062)	(0.063)	(0.062)	(0.062)	(0.062)	(0.059)	(0.063)	(0.057)	(0.065)	(0.061)	(0.063)
2	1.0	25	BIAS	-0.001	-0.001	-0.001	-0.002	-0.001	-0.001	0.003	-0.001	-0.028	-0.002	-0.009	-0.002
			ESE	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)	(0.055)	(0.053)	(0.053)	(0.052)	(0.053)	(0.052)	(0.053)
		50	BIAS	-0.002	-0.002	-0.001	-0.002	-0.002	-0.002	0.010	-0.001	-0.062	-0.003	-0.021	-0.002
			ESE	(0.059)	(0.060)	(0.062)	(0.060)	(0.060)	(0.060)	(0.059)	(0.059)	(0.056)	(0.058)	(0.058)	(0.058)
2	1.5	25	BIAS	0.001	0.000	0.001	0.000	0.001	0.001	-0.046	-0.004	-0.079	-0.007	-0.018	-0.003
			ESE	(0.089)	(0.088)	(0.090)	(0.088)	(0.088)	(0.088)	(0.083)	(0.087)	(0.081)	(0.087)	(0.085)	(0.086)
		50	BIAS	-0.002	-0.004	-0.001	-0.005	-0.003	-0.003	-0.101	-0.016	-0.174	-0.025	-0.048	-0.017
			ESE	(0.099)	(0.097)	(0.103)	(0.097)	(0.097)	(0.097)	(0.090)	(0.099)	(0.087)	(0.102)	(0.095)	(0.099)

Table 4: Markov model for event process and semi-Markov model for censoring process with $\lambda^c(t|H(t)) = [\lambda_0^c C^k]^{I(B(t)<\Delta)}$, $k = 1, 2, 3, 4$ and $\lambda^c(t|H(t)) = [\lambda_0^c C^4]^{I(B(t)<\Delta)}$, $k = 5, 6, \ldots$ ($\mu(\tau = 1) = 2$, $E = 1.5$, $C = 1.5$, $m = 500$, 1000 simulations)

Δ	% CENS		NA	WNA	PEPE	WPEPE	AJ	WAJ
0.1	0.25	BIAS	-0.0239	-0.0037	-0.0343	-0.0064	-0.0029	-0.0031
		ESE	(0.1055)	(0.1099)	(0.1031)	(0.1109)	(0.1093)	(0.1097)
	0.50	BIAS	-0.0488	-0.0154	-0.0611	-0.0160	-0.0211	-0.0204
		ESE	(0.1089)	(0.1206)	(0.1042)	(0.1423)	(0.1176)	(0.1185)
0.2	0.25	BIAS	-0.0447	-0.0064	-0.0677	-0.0097	-0.0045	-0.0047
		ESE	(0.1020)	(0.1083)	(0.0989)	(0.1091)	(0.1075)	(0.1081)
	0.50	BIAS	-0.0941	-0.0230	-0.1347	-0.0305	-0.0236	-0.0234
		ESE	(0.1076)	(0.1209)	(0.1023)	(0.1316)	(0.1187)	(0.1202)

λ_0, λ_0^d, E and D were specified so that $E\{N(1)\} = 1$ or 2, and the probability of the terminal event occuring before $\tau = 1$ was equal to 20% and 40% respectively. Given C, the censoring rate λ_0^c was determined to give a 25% or 50% rate of censoring at $\tau = 1$. Analyses were based on the unweighted and weighted Cook-Lawless estimator, unweighted and weighted Pepe estimators based on cumulative incidence functions, and unweighted and weighted Aalen-Johansen estimators. The general event-dependent censoring model in (3.6) was used in the analyses.

Semi-Markov models were also considered with $\lambda(t|H(t)) = \zeta_k(w)$ where $w = B(t)$ and $\zeta_k(w)$ is a cause-specific hazard function for the next event in state k. We adopted the hazard function of a gamma distribution with shape two as in Section 5.1 for this intensity, but retained the Markov intensities described above for death and determined β so that the expected number of events at $\tau = 1$ was one or two. The same censoring model and censoring rates were chosen as for the Markov process.

The empirical biases of the estimators of $\mu(t)$ are plotted over time in Figure 4 and indicate a considerable negative bias for the unweighted Pepe and unweighted Cook-Lawless estimators under the Markov process ($E = 1.25$, $D = 1.25$) with event-dependent censoring ($C = 1.5$); the weighted Pepe and Cook-Lawless estimators perform well. The unweighted Aalen-Johansen and weighted Aalen-Johansen (Datta-Satten) estimators perform well and are again indistinguishable in the Markov setting. In the semi-Markov setting ($\alpha = 2$, $D = 1.25$, $C = 1.5$) the Pepe estimator and the Aalen-Johansen estimator have large negative bias, but the Cook-Lawless estimator has small positive bias (comparable in magnitude to the Aalen-Johansen estimator). The weighted Pepe estimator still has some bias, but the weighted Cook-Lawless and weighted Aalen-Johansen (Datta-Satten) estimators perform much better.

The empirical biases and empirical standard errors in Tables 5 and 6 are for the estimators at $\tau = 1$ for the Markov and semi-Markov processes with terminal events respectively. The findings suggest all methods perform well with comparable efficiency under independent censoring for both Markov and semi-Markov processes. For Markov processes with dependent censoring, the empirical bias of the unweighted Aalen-Johansen estimator is negligible and the use of weighting has a remarkably small effect on the efficiency; the weighted Cook-Lawless and Pepe estimators again performed quite well. In the semi-Markov setting the weighted Cook-Lawless and weighted Aalen-Johansen (Datta-Satten) estimators perform very well in terms of bias and had comparable efficiencies.

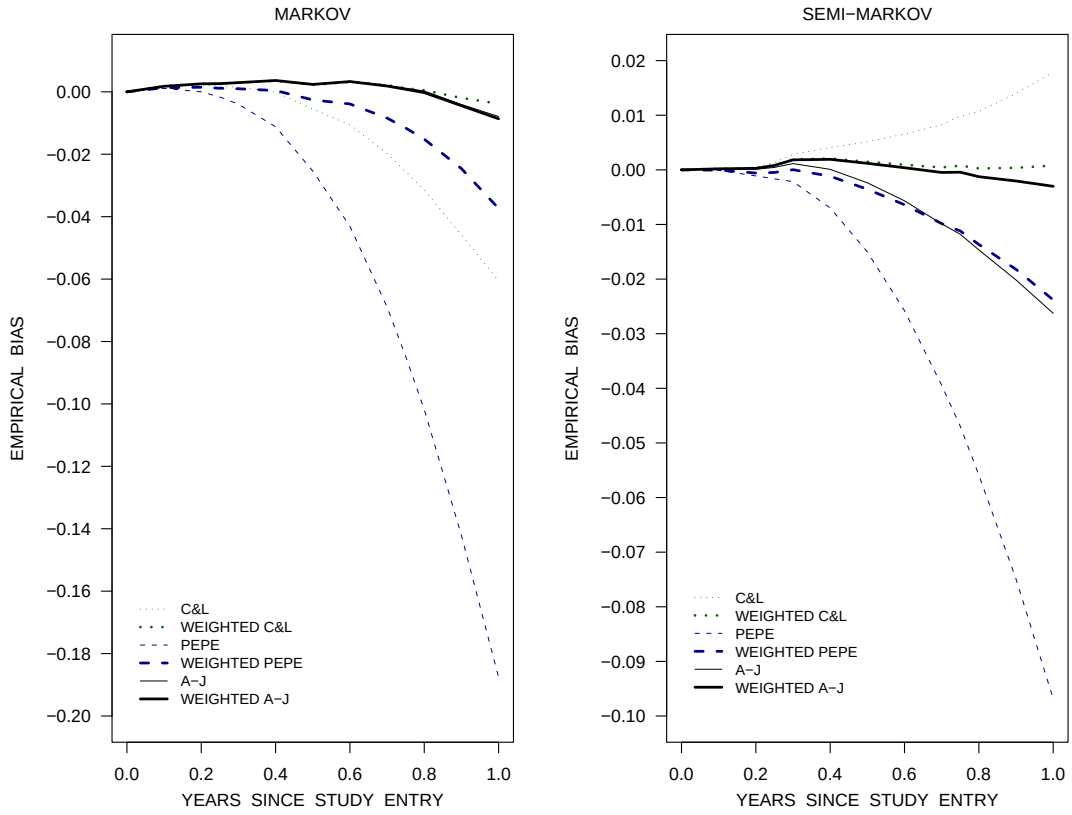

Figure 4: Empirical biases of unweighted and weighted estimates of mean function over time for Markov and semi-Markov processes with termination ($\mu(1) = 2$, 50% censoring at $\tau = 1$ with $C = 1.5$, $E = 1.25$ for Markov model and $\alpha = 2$ for semi-Markov model, probability terminal event occurs before $\tau = 1$ is 0.40 with $D = 1.25$).

Table 5: Empirical biases and standard errors for unweighted and weighted estimators under a Markov process with a terminal event (probability terminal event occurs before $\tau = 1$ is 0.40, $E = 1.25$, $\mu = E\{N(1)\}$, $m = 500$, 1000 simulations)

				C = 1.0						C = 1.5					
μ	D	% CENS		CL	WCL	PEPE	WPEPE	AJ	WAJ	CL	WCL	PEPE	WPEPE	AJ	WAJ
1	1.00	25	BIAS	0.003	0.003	-0.001	-0.002	0.003	0.003	-0.008	0.002	-0.030	-0.004	0.002	0.002
			ESE	(0.061)	(0.061)	(0.061)	(0.060)	(0.061)	(0.061)	(0.058)	(0.059)	(0.057)	(0.059)	(0.059)	(0.059)
		50	BIAS	0.004	0.002	-0.002	-0.005	0.003	0.003	-0.021	-0.001	-0.067	-0.012	-0.001	-0.001
			ESE	(0.070)	(0.070)	(0.071)	(0.069)	(0.070)	(0.070)	(0.065)	(0.070)	(0.061)	(0.069)	(0.070)	(0.069)
1	1.25	25	BIAS	0.003	0.002	-0.002	-0.002	0.003	0.002	-0.006	0.002	-0.027	-0.004	0.002	0.002
			ESE	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)	(0.056)	(0.057)	(0.054)	(0.056)	(0.057)	(0.057)
		50	BIAS	0.003	0.003	-0.002	-0.004	0.003	0.003	-0.015	0.001	-0.060	-0.009	0.000	0.000
			ESE	(0.067)	(0.066)	(0.068)	(0.066)	(0.066)	(0.066)	(0.065)	(0.069)	(0.061)	(0.069)	(0.069)	(0.069)
2	1.00	25	BIAS	0.005	0.004	-0.006	-0.007	0.004	0.004	-0.034	-0.000	-0.098	-0.017	-0.001	-0.001
			ESE	(0.102)	(0.101)	(0.102)	(0.101)	(0.101)	(0.101)	(0.098)	(0.102)	(0.093)	(0.101)	(0.101)	(0.102)
		50	BIAS	0.008	0.006	-0.005	-0.012	0.007	0.007	-0.081	-0.008	-0.213	-0.044	-0.013	-0.013
			ESE	(0.115)	(0.114)	(0.118)	(0.113)	(0.114)	(0.114)	(0.100)	(0.115)	(0.091)	(0.114)	(0.112)	(0.114)
2	1.25	25	BIAS	0.005	0.005	-0.004	-0.006	0.005	0.005	-0.025	0.002	-0.086	-0.014	0.001	0.001
			ESE	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.092)	(0.095)	(0.088)	(0.094)	(0.095)	(0.095)
		50	BIAS	0.008	0.006	-0.004	-0.010	0.007	0.007	-0.061	-0.004	-0.188	-0.037	-0.008	-0.009
			ESE	(0.109)	(0.108)	(0.111)	(0.108)	(0.108)	(0.108)	(0.099)	(0.111)	(0.090)	(0.109)	(0.108)	(0.110)

Table 6: Empirical biases and standard errors for unweighted and weighted estimators under a semi-Markov process (probability terminal event occurs before $\tau = 1$ is 0.40, $\alpha_E = 2.0$, $\mu = E\{N(1)\}$, $m = 500$, 1000 simulations)

				C = 1.0						C = 1.5					
μ	D	% CENS		CL	WCL	PEPE	WPEPE	AJ	WAJ	CL	WCL	PEPE	WPEPE	AJ	WAJ
1	1.00	25	BIAS	0.001	0.001	-0.005	-0.005	0.001	0.001	0.005	0.001	-0.017	-0.005	-0.003	0.001
			ESE	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.045)	(0.045)	(0.044)	(0.044)	(0.044)	(0.045)
		50	BIAS	0.001	0.001	-0.007	-0.008	0.001	0.001	0.010	0.001	-0.039	-0.008	-0.010	0.000
			ESE	(0.053)	(0.053)	(0.053)	(0.052)	(0.053)	(0.053)	(0.054)	(0.053)	(0.051)	(0.053)	(0.052)	(0.053)
1	1.25	25	BIAS	0.001	0.001	-0.004	-0.005	0.001	0.001	0.005	0.001	-0.016	-0.005	-0.003	0.001
			ESE	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.043)	(0.043)	(0.042)	(0.042)	(0.042)	(0.043)
		50	BIAS	0.001	0.002	-0.006	-0.007	0.001	0.001	0.012	0.002	-0.035	-0.007	-0.008	0.001
			ESE	(0.052)	(0.053)	(0.053)	(0.052)	(0.053)	(0.053)	(0.052)	(0.052)	(0.049)	(0.051)	(0.050)	(0.052)
2	1.00	25	BIAS	0.001	0.001	-0.016	-0.016	0.001	0.001	0.008	0.003	-0.048	-0.014	-0.009	0.002
			ESE	(0.067)	(0.067)	(0.067)	(0.066)	(0.067)	(0.067)	(0.066)	(0.066)	(0.064)	(0.065)	(0.065)	(0.066)
		50	BIAS	0.000	0.000	-0.024	-0.026	-0.000	-0.000	0.014	0.001	-0.107	-0.026	-0.029	-0.003
			ESE	(0.079)	(0.080)	(0.080)	(0.078)	(0.079)	(0.079)	(0.076)	(0.076)	(0.070)	(0.075)	(0.074)	(0.076)
2	1.25	25	BIAS	0.001	0.001	-0.015	-0.015	0.001	0.001	0.010	0.003	-0.044	-0.013	-0.008	0.002
			ESE	(0.063)	(0.063)	(0.063)	(0.062)	(0.063)	(0.063)	(0.063)	(0.063)	(0.061)	(0.062)	(0.062)	(0.063)
		50	BIAS	-0.000	-0.000	-0.023	-0.024	-0.001	-0.001	0.018	0.001	-0.097	-0.024	-0.026	-0.003
			ESE	(0.073)	(0.073)	(0.073)	(0.072)	(0.073)	(0.073)	(0.072)	(0.072)	(0.066)	(0.071)	(0.070)	(0.072)

6 SKELETAL COMPLICATIONS FROM BONE METASTASES

6.1 NONPARAMETRIC MEAN FUNCTION ESTIMATION

Here we apply the various methods discussed in Section 4 to estimate the expected number of skeletal events over time for each treatment arm of the study introduced in Section 1. Figure 5 contains plots of the Nelson-Aalen estimates of the cumulative transition rates for the occurrence of new skeletal events and death from states E_0, E_1, and E_2. These reveal increasing risk of new skeletal complications with the occurence of each skeletal event, particularly after the first event, and generally higher mortality rates with greater numbers of skeletal events. This is consistent with the notion that patients with a greater extent of metastatic disease have both greater risk of skeletal events and increased risk of death.

Figure 6 displays the Nelson-Aalen estimates of the cumulative transition rates for censoring according to the cumulative number of skeletal events (for those with 0, one, and two events) among patients in the control arm. Here the rate of censoring increases with each additional event indicating some event-dependent censoring. There is slightly different variation in the censoring rate in the pamidronate arm as well. This motivates the use of censoring models like (3.5). There may, however, be a more complex dependence of the censoring risk on the event history. Additional components of the event history were therefore introduced into the censoring intensity such as the time between the previous two events, with the effects of these times allowed to vary depending on the cumulative number of events. These analyses did not suggest the need to include additional terms and so a stratified model as in (3.6) was used.

The evidence of event-dependent censoring suggests that the Kaplan-Meier estimate of the survival function for time to death will be biased. Figure 7 contains plots of the unweighted and weighted Kaplan-Meier and Aalen-Johansen estimates for the cumulative distribution function of survival time. The six estimates of the mean function $\mu(t)$ are given in Figure 8, which reveals considerable differences in the placebo control arm. In the control arm, the three weighted estimates tend to agree more closely than the unweighted versions, with the Aalen-Johansen and Pepe estimates being at the two

extremes. For the pamidronate arm, four of the estimates agree reasonably well, but the unweighted Cook-Lawless and unweighted Pepe estimates are lower than the other estimates.

The marginal survivor functions and mean functions are easily interpreted and represent a natural basis for treatment comparisons for many questions of interest. This example illustrates the fact that failure to account for event-dependent censoring can lead to quite different estimates of the cumulative expected number of events. The simulations in Section 5 suggest there is little price to be paid for using inverse probability of censoring weights and so if diagnostic plots such as those in Figure 6 suggest there may be a form of event-dependent censoring, use of inverse weighting can substantially improve estimation. In the next section we discuss methods for estimating measures of treatment effect via semiparametric models.

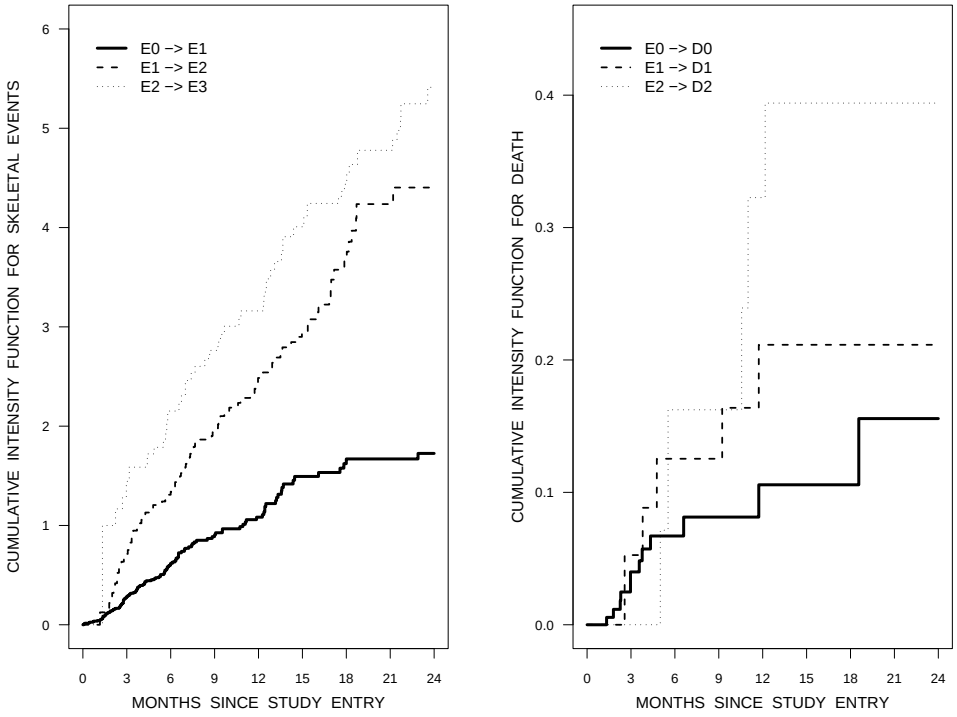

Figure 5: Nelson-Aalen estimates of the cumulative transition intensities for events and death for control patients in Hortobagyi et al. (1998)

6.2 Semiparametric Estimation of Treatment Effects

In addition to examining the effects of treatment on marginal features, we normally require quantitative specification, testing, and estimation of treatment effect. Here we describe how this can be done in the context of recurrent events and death, as represented by the bone metastases study.

Let $x_i = 1$ if individual i is randomized to the pamidronate arm and $x_i = 0$ if the individual receives the placebo. Here we consider a semiparametric model of the form $d\mu_i^*(t) = d\mu_0^*(t)\exp(\beta x_i)$ where $d\mu_0^*(t) = E\{dN_i(t)|T_i^d \geq t, x_i = 0\}$ is the rate of events at time t, given survival to t, for control subjects, and $\exp(\beta)$ is a multiplicative effect of treatment on this conditional rate. Let $\mathcal{F}_i^d(t) = P(T_i^d \geq t|x_i)$ denote the survivor function for subject i with covariate x_i. A proportional hazards model $dH_i^d(t) = dH_0^d(t)\exp(\gamma x_i)$ can also be formed, giving $\mathcal{F}_i^d(t) = \exp(-H_0^d(t)\exp(\gamma x_i))$. Note that $\int_0^t d\mu_i^*(u)$ is uninterpretable, but $\int_0^t \mathcal{F}_i^d(u)d\mu_i^*(u)$ is the expected number of events for subject i over $(0, t]$, accounting for the possibility of termination; see (4.1). This semiparametric formulation

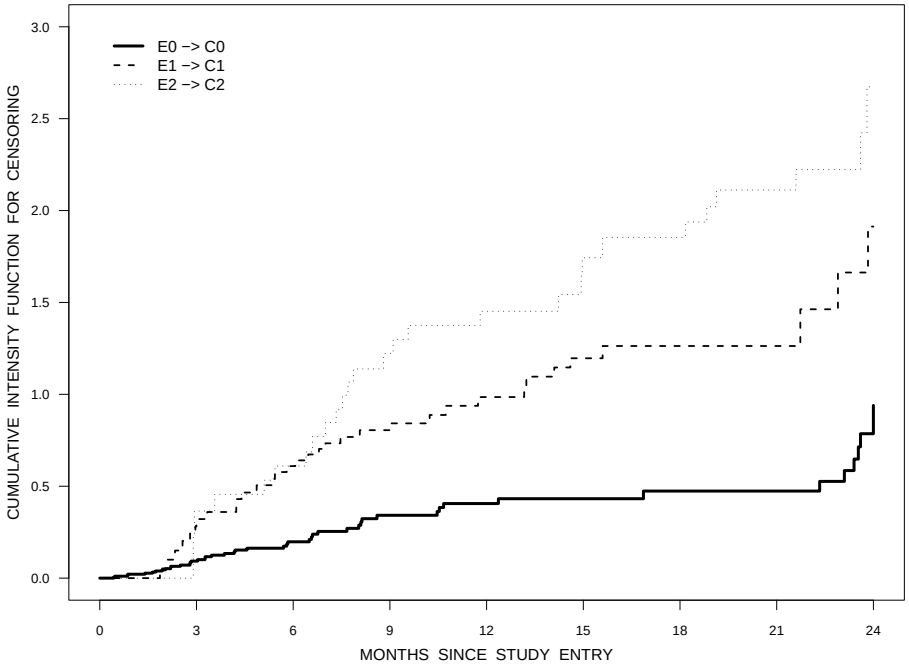

Figure 6: Nelson-Aalen estimates of the cumulative intensities for censoring for control patients in Hortobagyi et al. (1998)

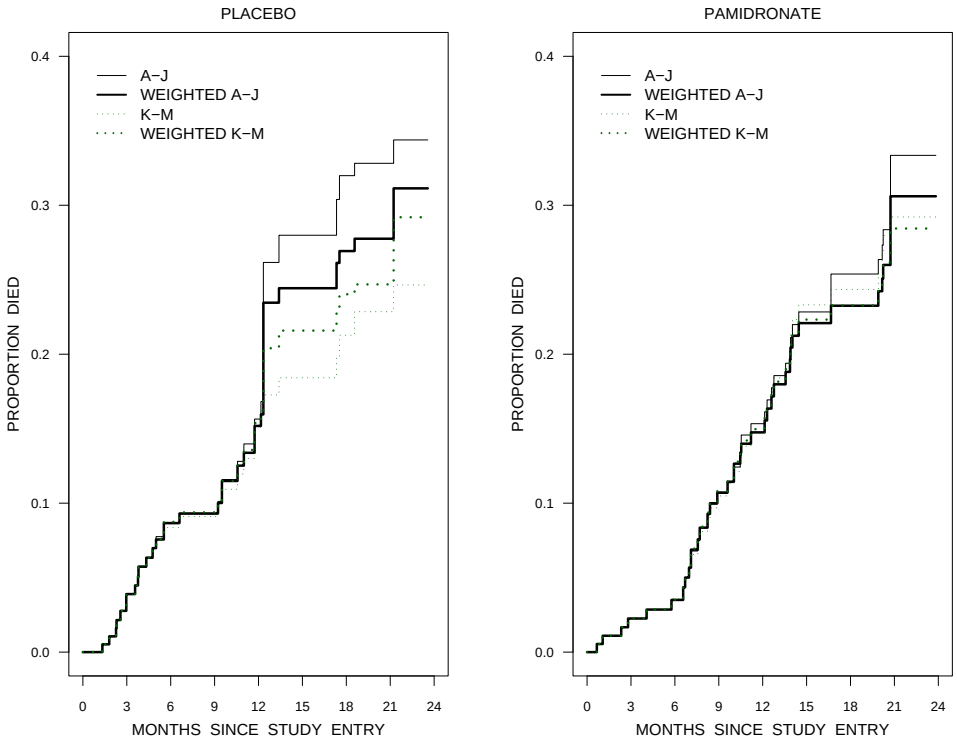

Figure 7: Various estimates of the cumulative probability of death for control patients in Hortobagyi et al. (1998)

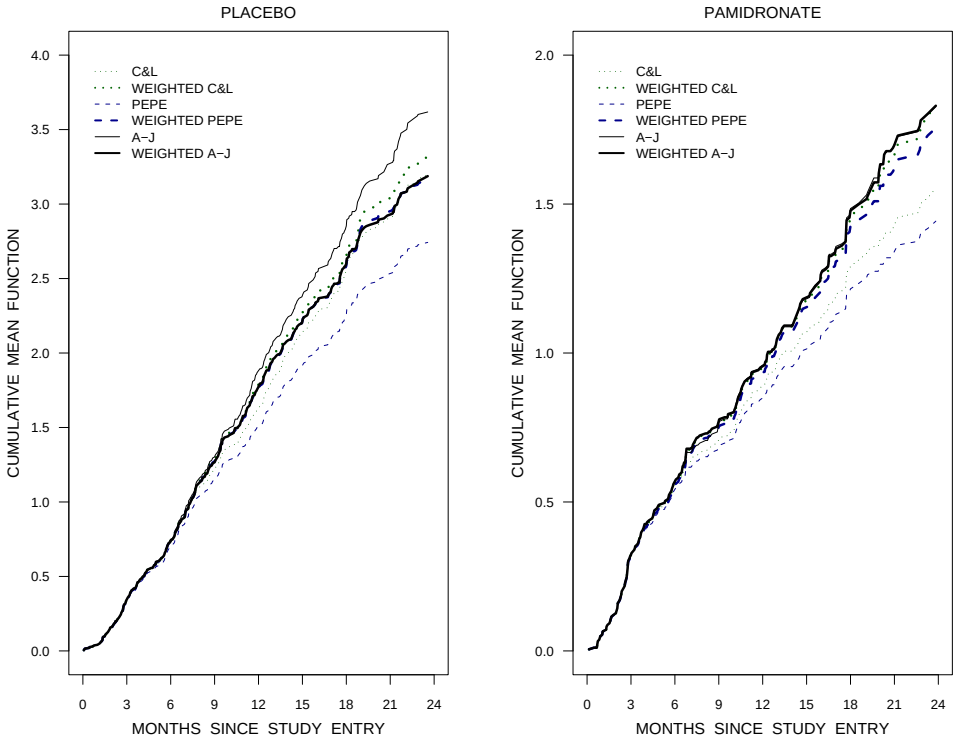

Figure 8: Unweighted and weighted estimates of the cumulative mean function for treated (pamidronate) and control (placebo) patients in Hortobagyi et al. (1998)

allows one to obtain simple estimates of treatment effects; note that the plots described in Section 6.1 allow us to assess the adequacy of the multiplicative models.

To protect against event-dependent censoring, we consider IPCW estimating equations for the conditional rate model, of the form

$$U_1(t;\beta) = \sum_{i=1}^{m} \frac{C_i(t)Y_i^d(t)}{\widehat{G}_i(t)}[dN_i(t) - \exp(\beta x_i)d\mu_0^*(t)] = 0\,, \quad t > 0 \tag{6.1}$$

$$U_2(\tau;\beta) = \sum_{i=1}^{m} \int_0^\tau \frac{C_i(t)Y_i^d(t)x_i}{\widehat{G}_i(t)}[dN_i(t) - \exp(\beta x_i)d\mu_0^*(t)] = 0\,. \tag{6.2}$$

Solving $U_1(t;\beta) = 0$ gives

$$d\tilde{\mu}_0^*(t;\beta) = \frac{\sum_{i=1}^{m} C_i(t)Y_i^d(t)dN_i(t)/\widehat{G}_i(t)}{\sum_{i=1}^{m} C_i(t)Y_i^d(t)\exp(\beta x_i)/\widehat{G}_i(t)}$$

which can in turn be substituted into (6.2) to obtain

$$\tilde{U}_2(\tau;\beta) = \sum_{i=1}^{m} \int_0^\tau \frac{C_i(t)Y_i^d(t)x_i}{\widehat{G}_i(t)} \left\{ dN_i(t) - \exp(\beta x_i) \frac{\sum_{\ell=1}^{m} C_\ell(t)Y_\ell^d(t)dN_\ell(t)/\widehat{G}_\ell(t)}{\sum_{\ell=1}^{m} C_\ell(t)Y_\ell^d(t)\exp(\beta x_\ell)/\widehat{G}_\ell(t)} \right\}, \tag{6.3}$$

which can be set equal to zero and solved to give $\widehat{\beta}$.

Analogous equations can be defined to obtain $\widehat{\gamma}$ as described in Satten et al. (2001). The IPCW estimating equations for the failure time model are of the form

$$U_1^d(t;\gamma) = \sum_{i=1}^{m} \frac{C_i(t)Y_i^d(t)}{\widehat{G}_i(t)}[dN_i^d(t) - \exp(\gamma x_i)dH_0^d(t)] = 0\,, \quad t > 0 \tag{6.4}$$

$$U_2^d(\tau;\gamma) = \sum_{i=1}^{m} \int_0^\tau \frac{C_i(t)Y_i^d(t)x_i}{\widehat{G}_i(t)}[dN_i^d(t) - \exp(\gamma x_i)dH_0^d(t)] = 0\,. \tag{6.5}$$

Solving $U_1^d(t;\gamma) = 0$ gives

$$d\tilde{H}_0^d(t;\gamma) = \frac{\sum_{i=1}^{m} C_i(t)Y_i^d(t)dN_i^d(t)/\widehat{G}_i(t)}{\sum_{i=1}^{m} C_i(t)Y_i^d(t)\exp(\gamma x_i)/\widehat{G}_i(t)}$$

which can in turn be substituted into (6.5) to give an equation analogous to (6.3),

$$\tilde{U}_2^d(\tau;\gamma) = \sum_{i=1}^{m} \int_0^\tau \frac{C_i(t)Y_i^d(t)x_i}{\widehat{G}_i(t)} \left\{ dN_i^d(t) - \exp(\gamma x_i) \frac{\sum_{\ell=1}^{m} C_\ell(t)Y_\ell^d(t)dN_\ell^d(t)/\widehat{G}_\ell(t)}{\sum_{\ell=1}^{m} C_\ell(t)Y_\ell^d(t)\exp(\gamma x_\ell)/\widehat{G}_\ell(t)} \right\} \tag{6.6}$$

which can be solved to give $\widehat{\gamma}$. Variance estimation is rather complicated for $(\widehat{\beta}, \widehat{\gamma})$ and we use a nonparametric bootstrap to compute the standard errors that follow.

In attempting to obtain estimates of β and γ for the bone metastases data, at some of the later time points the estimates of $G_i(t)$ became close to zero. In what follows we therefore use data over the first 18 months of follow-up. An alternative is to use a model that specifies the effect of $N_i(t^-)$ parametrically instead of (3.6).

If an unweighted analysis is carried out using the data over the first 18 months of follow-up, we find $\widehat{\beta} = -0.617$ (s.e.($\widehat{\beta}$)= 0.095, RR= $\exp(-0.617) = 0.540$, p<0.0001); a weighted analysis gives a slightly smaller estimate $\widehat{\beta} = -0.584$ (s.e.($\widehat{\beta}$)= 0.182, RR= $\exp(-0.584) = 0.558$, p=0.0013); both are highly significant, and suggest a roughly 45% reduction in the rate of skeletal events. Note that the ratio of the treatment to control mean functions in Figure 8 is approximately 0.50, which is reasonably close to these estimates. It is important to note, however, that this is estimating something slightly different than the multiplicative effect obtained from (6.1) and (6.2); the estimates in Figure 8 are marginal and allow for death, whereas the approach here considers event rates conditional on being alive.

In the analysis of the treatment effect on the time to death, we obtain $\widehat{\gamma} = 0.085$ (s.e.($\widehat{\beta}$)= 0.266, p=0.7502) for an unweighted analysis and $\widehat{\gamma} = -0.028$ (s.e.($\widehat{\gamma}$)= 0.268, p=0.9178) for the weighted analysis, both being compatible with no treatment effect on the time to death.

7 DISCUSSION

This article demonstrates that dependent censoring may arise even in carefully designed phase III clinical trials. This can cause biases when interest lies in marginal features of recurrent event or multistate life history processes (e.g., Klein et al., 2000). Estimates of marginal features can typically be based on simple estimating functions or on more complex models, from which estimates of marginal

features can be obtained. The Nelson-Aalen and Cook-Lawless estimates of $\mu(t)$ are examples of simple marginal estimates without and with terminal events, respectively. The Aalen-Johansen estimate is an example of the more complex modeling approach, which also leads to an estimate of the mean function. In addition, however, it provides estimates of marginal state occupancy probabilities, which convey a more complete picture of the distribution of event counts over time.

There is some built-in protection against dependent censoring with intensity-based or partly conditional models, because the censoring is only required to be conditionally independent as discussed in Section 1.2. However, this protection comes from specifying the model more completely and depends on the adequacy of the model used, as we saw here for the Aalen-Johansen estimator.

Inverse weighting offers an appealing alternative approach; it allows the use of simple marginal estimates and seems to more or less eliminate bias due to event-dependent censoring, with little price in terms of efficiency. In some of the the simulations we conducted, we used a correctly specified censoring process, but in practice the censoring mechanism is unknown and careful diagnostic checks of the censoring process and its dependence on prior event history or covariates are called for. We note that (1) limited simulations have shown some robustness of IPCW estimators to misspecification of the censoring intensity, and (2) that good diagnostic methods exist for assessing assumptions about censoring and other time-to-event hazard functions (e.g., Kalbfleisch and Prentice, 2002; Lawless, 2003). We note also that certain authors have examined additive models for censoring in connection with the estimation of state occupancy probabilities by Aalen-Johansen and weighted Aalen-Johansen estimators, for general multistate models (e.g., Datta and Satten, 2002; Gunnes et al., 2007). Such models can also be valuable in our setting and, in fact, (3.5) can be considered such a model.

We have not discussed variance estimation. Glidden (2002) developed robust variance estimates for the Aalen-Johansen estimates of the $P_{0k}(0,t)$ in Section 3.3, which could be used to obtain robust standard errors for the Aalen-Johansen estimates of the mean function. Datta and Satten (2002) develop variance estimates for the Datta-Satten estimator of the state occupancy probabilities, which could be programmed to obtain confidence intervals for the mean function. They are fairly complicated expressions and we recommend using resampling to obtain variance estimates.

Finally, we note the broad appeal of weighted Nelson-Aalen estimates for estimation of mean functions; they tend to perform the best across the situations we have examined, and are simple to implement. However, if there is a desire to estimate the distribution of times to some specific event, or the distribution of the number of recurrent events up to time t, then the weighted Pepe and Aalen-Johansen approaches are useful, and also provide good estimates of mean functions.

ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (RJC, JFL) and the Canadian Institutes for Health Research (RJC). R.J. Cook is a Canada Research Chair in Statistical Methods for Health Research. The authors thank Novartis Pharmaceuticals for permission to use the data from the pamidronate trial.

REFERENCES

Aalen, O.O., Borgan, O., and Fekjaer, H. (2001). "Covariate Adjustment of Event Histories Estimated with Markov Chains: The Additive Approach," *Biometrics*, 57, 993-1001.

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

Andersen, P.K. and Gill, R.D. (1982). "Cox's Regression Model for Counting Processes: A Large Sample Study," *Annals of Statistics*, 10, 1100-1120.

Cook, R.J. and Lawless, J.F. (1997). "Marginal Analysis of Recurrent Events and a Terminal Event," *Statistics in Medicine*, 16, 911-924.

Cook, R.J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*, New York: Springer.

Couper, D. and Pepe, M.S. (1997). "Modelling Prevalence of a Condition: Chronic Graft-Versus-Host Disease After Bone Marrow Transplantation," *Statistics in Medicine*, 16, 1551-1571.

Cox, D.R. and Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events*, London: Chapman and Hall,

Cox, D.R. and Miller, H.D. (1965). *The Theory of Stochastic Processes*. London: Chapman and Hall/CRC.

Crowder, M.J. (2001). *Classical Competing Risks*. London: Chapman and Hall/CRC,

Datta, S. and Satten, G.A. (2001). "Validity of the Aalen-Johansen Estimators of Stage Occupation Probabilities and Nelson-Aalen Estimators of Integrated Transition Hazards for non-Markov Models," *Statistics and Probability Letters*, 55, 403-411.

Datta, S. and Satten, G.A. (2002). "Estimation of Integrated Transition Hazards and Stage Occupation Probabilities for non-Markov Systems Under Dependent Censoring," *Biometrics*, 58, 792-802.

Duchateau, L., Janssen, P., Kezic, I., and Fortpied, C. (2003). "Evolution of Recurrent Asthma Event Rate Over Time in Frailty Models," *Applied Statistics*, 52, 355-363.

Ghosh, D. and Lin, D.Y. (2000). "Nonparametric Analysis of Recurrent Events and Death," *Biometrics*, 56, 554-562.

Ghosh, D. and Lin, D.Y. (2002). "Marginal Regression Models for Recurrent and Terminal Events," *Statistica Sinica*, 12, 663-688.

Glidden, D. (2002). "Robust Inference for Event Probabilities with non-Markov Event Data," *Biometrics*, 58, 361-368.

Gunnes, N., Borgan, O. and Aalen, O.O. (2007). "Estimating Stage Occupation Probabilities in non-Markov Models," *Lifetime Data Analysis*, 13, 211-240.

Hortobagyi, G.N., Thierault, R.L., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J.F., Seaman, J., Knight, R.D., Heffernan, M., and Reitsma, D.J. (1996). "Efficacy of Pamidronate in Reducing Skeletal Complications in Patients with Breast Cancer and Lytic Bone Metastases," *New England Journal of Medicine*, 335, 1785-1791.

Hortobagyi, G.N., Theriault, R.L., Lipton, A., Porter, L., Blayney, D., Sinoff, C., Wheeler, H., Simeone, J.F., Seaman, J., Knight, R.D., Heffernan, M., Mellars, K., and Reitsma, D.J. (1998). Long-term Prevention of Skeletal Complications of Metastic Breast Cancer with Pamidronate," *Journal of Clinical Oncology*, 16, 2038-2044.

Huang, C.Y. and Wang, M.C. (2004). "Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data," *Journal of the American Statistical Association*, 99, 1153-1165.

Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Kanis, J.A., Powles, T., Paterson, A.H.G., McCloskey, E.V., and Ashley, S. (1996). "Clodronate Decreases the Frequency of Skeletal Metastases in Women with Breast Cancer," *Bone*, 19, 663-667.

Klein, J.P., Keiding, N., Shu, Y., Szydlo, R.M., and Goldman, J.M. (2000). "Summary Curves for Patients Transplanted for Chronic Myeloid Leukaemia Salvaged by a Donor Lymphocyte Infusion: the Current Leukaemia-free Survival Curve," *British Journal of Haematology*, 109, 148-152.

Lawless, J.F. (1987a). "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association*, 82, 808-815.

Lawless, J.F. (1987b). "Negative Binomial and Mixed Poisson Regression," *The Canadian Journal of Statistics*, 15, 209-225.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd Edition), Hoboken, N.J: John Wiley and Sons.

Lawless, J.F., and Nadeau, J.C. (1995). "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158-168.

Lawless, J.F., Nadeau, J.C., and Cook, R.J. (1997). "Analysis of Mean and Rate Functions for Recurrent Events," *In Proceedings of the First Seatle Survival Analysis Symposium*, D.Y. Lin and T.R. Fleming (eds), 137-140. New York:Springer.

Lin, D.Y., Sun, W., and Ying, Z. (1999). "Nonparametric Estimation of the Gap Time Distribution for Serial Events with Censored Data," *Biometrika*, 86, 59-70.

Lin, D.Y., Wei, L.J., Yang, I., and Ying, Z. (2000). "Semiparametric regression for the mean and rate functions of recurrent events," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 711-731.

Liu, L., Wolfe, R.A., and Huang, X. (2004). "Shared Frailty Models for Recurrent Events and a Terminal Event," *Biometrics*, 60, 747-756.

Miloslavsky, M., Keles, S., van der Laan, M.J., and Butler, S. (2004). "Recurrent Events Analysis in the Presence of Time-dependent Covariates and Dependent Censoring," *Journal of the Royal Statistical Society, Series B*, 66, 239-257.

Moecks, J., Koch, G.G., Scott, M., Budde, M., Givens, S., Koehler, W., and Knoessl, M. (2004). "Measures of Morbidity in Clinical Studies with Recurrent Skeletal Complications," *Journal of Biopharmaceutical Statistics*, 14, 415-437.

Pepe, M.S. (1991). "Inference for Events with Dependent Risks in Multiple Endpoint Studies," *Journal of the American Statistical Association*, 86, 770-778.

Pepe, M.S., Longton, G., and Thornquist, M. (1991). "A Qualifier Q for the Survival Function to Describe the Prevalence of a Transient Condition," *Statistics in Medicine*, 10, 413-421.

Pepe, M.S. and Cai J. (1993). "Some Graphical Displays and Marginal Regression Analyses for Recurrent Failure Times and Time Dependent Covariates," *Journal of the American Statistical Association* 88, 811-820.

Robins, J.M. (1993). "Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers," In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, Alexandria, Virginia: American Statistical Association, 24-33.

Robins, J.M. and Rotnitzky, A. (1992). "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," In *AIDS Epidemiology-Methodological Issues*, 279–331. Eds. N. Jewell, K. Dietz, and V. Farewell. Boston: Birkäuser.

Robins, J.M. and Rotnitzky, A. (1995). "Semiparametric Regression Estimation in the Presence of Dependent Censoring," *Biometrika*, 82, 805-820.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., and Soubeyran, P. (2007). "Joint Frailty Models for Recurring Events and Death Using Maximum Penalized Likelihood Estimation: Application on Cancer Events," *Biostatistics*, 8, 708-721.

Satten, G.A., Datta, S., and Robins J.M. (2001). "An Estimator for the Survival Function When Data are Subject to Dependent Censoring," *Statistics and Probability Letters*, 54, 397-403.

Schaubel, D.E. and Cai, J. (2004). "Regression Methods for Gap Time Hazard Functions of Sequentially Ordered Multivariate Failure Time Data," *Biometrika*, 91, 291-303.

Thall, P.F. and Vail, S.C. (1990). "Some Covariance Models for Longitudinal Count Data with Overdispersion," *Biometrics*, 46, 657-671.

Therneau, T.M. and Hamilton, S.A. (1997). "rhDNase as an Example of Recurrent Event Analysis," *Statistics in Medicine*, 16, 2029-2047.

van der Laan, M. Hubbard, A., and Robins, J. (2002). "Locally Efficient Estimation of a Multivariate Survivor Function in Longitudinal Studies," *Journal of the American Statistical Association*, 98, 494-507.

Wei, L.J. and Glidden, D.V. (1997). "An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials," *Statistics in Medicine*, 16, 833–839.

Wei, L.J., Lin, D.Y., and Weissfeld, L. (1989). "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *Journal of the American Statistical Association*, 84, 1065-1073.