Semiparametric Inference for a General Class of Models for Recurrent Events

E. A. Peña* E. H. Slate† J. R. González‡

December 10, 2003

Abstract

Procedures for estimating the parameters of the general class of semiparametric models for recurrent events proposed by Peña and Hollander (2004) are developed. This class of models incorporates an effective age function which encodes the changes that occur after each event occurrence such as the impact of an intervention, it allows for the modeling of the impact of accumulating event occurrences on the unit, it admits a link function in which the effect of possibly time-dependent covariates are incorporated, and it allows the incorporation of unobservable frailty components which induce dependencies among the inter-event times for each unit. The estimation procedures are semiparametric in that a baseline hazard function is nonparametrically specified. The sampling distribution properties of the estimators are examined through a simulation study, and the consequences of mis-specifying the model are analyzed. The results indicate that the flexibility of this general class of models provides a safeguard for analyzing recurrent event data, even data possibly arising from a frailty-less mechanism. The estimation procedures are applied to real data sets arising in the biomedical and public health settings, as well as from reliability and engineering situations. In particular, the procedures are applied to a data set pertaining to times to recurrence of bladder cancer and the results of the analysis are compared to those obtained using three methods of analyzing recurrent event data.

Key Words and Phrases: Correlated inter-event times; counting process; effective age process; EM algorithm; frailty; intensity models; model mis-specification; sum-quota accrual scheme.

1 Introduction

Recurrent events occur in many settings such as in biomedicine, public health, clinical trials, engineering and reliability studies, politics, economics, sociology, actuary, among others. Examples

*E. Peña is Professor in the Department of Statistics, University of South Carolina, Columbia, SC 29208. His e-mail address is pena@stat.sc.edu. He acknowledges the research support provided by NSF Grant DMS 0102870, NIH Grant GM056182, NIH COBRE Grant RR17698, and the USC/MUSC Collaborative Research Program.

†E. Slate is Associate Professor in the Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, SC 29425. Her e-mail address is slateeh@musc.edu. She acknowledges the research support provided by NIH Grant CA077789, NIH COBRE Grant RR17696, DAMD Grant 17-02-1-0138 and the USC/MUSC Collaborative Research Program.

‡J González is Associate Professor in the Department of Biostatistics, International University of Catalonia, and Associate Researcher in the Cancer Prevention and Control Unit, Catalan Institute of Oncology, Avda. Gran via s/n, km. 2.7, Hospitalet de Llobregat 08907, Spain. His e-mail address is jrgonzalez@iconcologia.catsalut.es. He acknowledges the research support provided by co-operative Research Networks in Cancer (Contract C03/10) and Epidemiology and Public Health (Contract C03/09), funded by the Carlos III Institute, Spanish Ministry of Health.

of recurrent events in the biomedical and public health settings are the re-occurrence of a tumor after surgical removal in cancer studies, epileptic seizures, drug or alcohol abuse of adolescents, outbreak of a disease such as encephalitis, recurring migraine headaches, hospitalization, movement in the small bowel during fasting state, onset of depression, nauseous feeling when taking drugs for the dissolution of cholesterol gallstones, recurrence of caries, ulcers or inflammation in an oral health study, and angina pectoris for patients with coronary disease. Some other specific biomedical examples of recurrent events are described in Cook and Lawless (2002). In the engineering and reliability settings, recurrent events could be the breakdown or failure of a mechanical or electronic system, the discovery of a bug in an operating system software, the occurrence of a crack in concrete structures, the breakdown of a fiber in fibrous composites, among others. Non-life insurance claims, traffic accidents, terrorist attacks, the Dow Jones Industrial Average decreasing by more than 200 points on a trading day, change of employment, among many others, are but a few examples of recurrent phenomena in other settings.

There are currently several models and methods of analysis used for recurrent event data. See for example Hougaard (2000), Therneau and Hamilton (1997), and Therneau and Grambsch (2000) for some current approaches to analyzing recurrent event data. However, as pointed out in Peña and Hollander (2004), there is still a need for a general and flexible class of models that *simultaneously* incorporates the effects of covariates or concomitant variables, the impact on the unit of accumulating event occurrences, the effect of performed interventions after each event occurrence, as well as the effect of latent or unobserved variables which, for each unit, endow correlation among the inter-event times. In recognition of this need, Peña and Hollander (2004) proposed a general class of models for recurrent events which satisfies the above requirements. This class of models will be described in Section 2. The current paper deals with inference issues, specifically the estimation of parameters, for this new class of models. However, we limit the scope of this paper to examining the finite-sample properties through simulation studies of the resulting estimators and defer the analytical and asymptotic analysis of their properties to a forthcoming paper.

For our setting, we consider an observational unit (e.g., a patient in a biomedical setting, an electronic system in a reliability setting) which is being monitored for the occurrence of a recurrent event over a study period $[0, \tau]$, where τ may represent an administrative time, time of study termination, or some other right-censoring variable. The time τ could be a random time governed by an unknown probability distribution function $G(t) = \mathbf{P}(\tau \le t)$. Let $S_0 \equiv 0 < S_1 < S_2 < S_3 < \dots$ be the successive calendar times of event occurrences, and let $T_1, T_2, T_3, \dots$ be the times between successive event occurrences. Thus, for $i = 1, 2, 3, \dots$, $T_i = S_i - S_{i-1}$ and $S_i = T_1 + T_2 + \dots + T_i$. Over the observation period $[0, \tau]$, the number of event occurrences is $K = \max\{k \in \{0, 1, 2, \dots\} : S_k \le \tau\}$, which is a random variable whose distribution depends on the distributional properties of the inter-occurrence times T_is and the distribution G of τ. As such, K is *informative* with regards to the distributional properties of event occurrences.

Assume also that for this unit there is a, possibly time-varying, q-dimensional vector of covariates such as gender, age, race, disease status, white blood cell counts (WBC), prostate specific antigen (PSA) level, weight, blood pressure, treatment regimen, etc. We suppose that over the period $[0, \tau]$, the realization of this covariate process is observable. We denote this covariate process by $\left\{\mathbf{X}(s) = (X_1(s), X_2(s), \dots, X_q(s))^{\mathrm{t}} : 0 \le s \le \tau\right\}$, with "t" representing vector/matrix transpose. For this subject, the observable entities over the study period $[0, \tau]$ are therefore

$$\mathbf{D}(\tau) \equiv \left\{(\mathbf{X}(s) : 0 \le s \le \tau), K, \tau, T_1, T_2, \dots, T_K, \tau - S_K\right\}. \tag{1}$$

Notice that since $S_K = \sum_{j=1}^{K} T_j$, specifying the value of $\tau - S_K$ renders specifying τ redundant; however, we still include this to indicate that $\tau - S_K$ is the right-censoring variable for the inter-occurrence time T_{K+1}. Furthermore, note that since K is a random variable, then the distributional properties of both $\tau - S_K$ and T_{K+1} maybe of a complicated form. We remark that when considering the data structure in this recurrent event situation, there is a need to recognize that K is informative and that the censoring mechanism for T_{K+1} is informative (cf., Wang and Chang 1999; Lin, Sun, and Ying 1999; Peña, Strawderman, and Hollander 2001). These aspects are borne out of the *sum-quota* data accrual scheme, since the total number of observed events is intrinsically tied to the distributions governing the event occurrences themselves.

The observable entities may also be represented more succinctly and more beneficially through the use of stochastic processes. Still considering one unit, let us define for calendar time s, $N^\dagger(s) = \sum_{j=1}^{\infty} I\{S_j \le s, S_j \le \tau\}$, where $I\{\cdot\}$ denotes indicator function, the process which counts the number of events observed on or before calendar time s during the study period $[0, \tau]$. Furthermore, define for calendar time s, $Y^\dagger(s) = I\{\tau \ge s\}$, the "at-risk" process which indicates whether the subject is still under observation at calendar time s or not. The data $\mathbf{D}(\tau)$ in (1) could be represented by

$$\mathbf{D}(s^*) = \left\{ \left(\mathbf{X}(s), N^\dagger(s), Y^\dagger(s) \right) : 0 \le s \le s^* < \infty \right\}, \tag{2}$$

where s^* is an upper limit of observation time. Note that even though $\mathbf{X}(s)$ is not observed for $s > \tau$ this does not pose a problem since for $s > \tau$, $Y^\dagger(s) = 0$, and so for such a subject, there will be no more information obtainable past τ. If in the study there are n subjects, the observables will be $\mathbf{D}(s^*) = (\mathbf{D}_1(s^*), \mathbf{D}_2(s^*), \ldots, \mathbf{D}_n(s^*))$, where for $i = 1, 2, \ldots, n$,

$$\mathbf{D}_i(s^*) = \left\{ (\mathbf{X}_i(s) : s \le \tau_i), K_i, \tau_i, T_{i1}, T_{i2}, \ldots, T_{iK_i}, \tau_i - S_{iK_i} \right\}. \tag{3}$$

Equivalently, $\mathbf{D}_i(s^*) = \left\{ \left(\mathbf{X}_i(s), N_i^\dagger(s), Y_i^\dagger(s) \right) : 0 \le s \le s^* < \infty \right\}$, where $N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \le s, S_{ij} \le \tau_i\}$ and $Y_i^\dagger(s) = I\{\tau_i \ge s\}$, and $S_{i1} < S_{i2} < \ldots$ are the calendar times of successive event occurrences for the ith subject, $T_{ij} = S_{ij} - S_{ij-1}, j = 1, 2, \ldots$, and τ_i is the censoring time of the ith subject.

Before proceeding, we briefly provide an outline of the contents of this paper. As mentioned earlier, Section 2 will present a description of the class of models for recurrent events that is under investigation. Section 3 will examine the problem of estimating the parameters of the model when there are no frailty components. The results here are needed for dealing with the case with frailties, so Section 4 describes the estimation procedure in the presence of frailties. Section 5 will summarize results of the simulation studies pertaining to the properties of the estimators. We demonstrate the estimation procedures discussed in Sections 3 and 4 on real data sets in Section 6. Section 7 will provide some concluding thoughts.

2 A General Class of Models

In this section we describe the general class of models for recurrent events proposed in Peña and Hollander (2004). Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ be a vector of independent and identically distributed (i.i.d.) positive-valued random variables from a parametric distribution $H(z;\xi) = \mathbf{P}(Z \leq z|\xi)$ where ξ is a finite-dimensional parameter taking values in $\Xi \subseteq \Re^r$. These variables are unobservable random factors affecting the event occurrences for the subjects. Also, let $\mathbf{F} = \{\mathcal{F}_s : 0 \leq s \leq s^*\}$ be a filtration or history on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{X}_i$s and $Y_i^\dagger$s are predictable and such that $N_i^\dagger$s are counting processes with respect to $\mathbf{F}$. The general class of models requires the specification, possibly done dynamically, of predictable observable processes $\{\mathcal{E}_i(s) : 0 \leq s \leq s^*\}$, $i = 1, 2, \ldots, n$, satisfying the following conditions: (I) $\mathcal{E}_i(0) = e_{i0}$, almost surely (a.s.), where $e_{i0}, i = 1, 2, \ldots, n$, are nonnegative real numbers; (II) $\mathcal{E}_i(s) \geq 0, i = 1, 2, \ldots, n$; and (III) On $[S_{ik-1}, S_{ik})$, $\mathcal{E}_i(s)$ is monotone and almost surely differentiable with a positive derivative $\mathcal{E}_i'(s)$. The class of models is obtained by postulating that, conditionally on $\mathbf{Z}$, the $\mathbf{F}$-compensator of $N_i^\dagger$ is $\{A_i^\dagger(s|\mathbf{Z}, \mathbf{X}_i) : 0 \leq s \leq s^*\}$ with

$$A_i^\dagger(s|\mathbf{Z}, \mathbf{X}_i) = \int_0^s Y_i^\dagger(v)\, \lambda_i(v|\mathbf{Z}, \mathbf{X}_i)\, \mathrm{d}v; \tag{4}$$

$$\lambda_i(s|\mathbf{Z}, \mathbf{X}_i) = Z_i\, \lambda_0[\mathcal{E}_i(s)]\, \rho[N_i^\dagger(s-); \alpha]\, \psi[\beta^{\mathrm{t}} \mathbf{X}_i(s)]. \tag{5}$$

This means that the process $M_i^\dagger(s|\mathbf{Z}, \mathbf{X}_i) = N_i^\dagger(s) - A_i^\dagger(s|\mathbf{Z}, \mathbf{X}_i)$ is a square-integrable $\mathbf{F}$-martingale. In (5), $\lambda_0(\cdot)$ is an unknown baseline hazard rate function; $\rho(\cdot;\alpha) : \mathbf{Z}_+ \equiv \{0, 1, 2, \ldots\} \to \Re_+$ is of known functional form with $\rho(0;\alpha) = 1$ and with $\alpha \in \mathcal{A} \subseteq \Re^p$; and $\psi(\cdot)$ is a nonnegative link function of known functional form with $\beta \in \mathcal{B} \subseteq \Re^q$. The unknown model parameters are $(\lambda_0(\cdot), \alpha, \beta, \xi)$, where $\lambda_0(\cdot)$ is non-parametrically specified, and α, β, and ξ are finite-dimensional parameters.

The main impetus in introducing this general class of models for recurrent events is that it incorporates simultaneously the effects of covariates through the link function $\psi(\cdot)$, the associations among the event inter-occurrence times through the unobservable frailty variables Z_is, the effects attributable to the accumulating event occurrences for a subject through the component $\rho(\cdot;\alpha)$, and the effects of performed interventions after each event occurrence through the effective age processes $\{\mathcal{E}_i(s)\}$ that act on the baseline hazard rate function $\lambda_0(\cdot)$. By requiring condition (III) together with $\mathcal{E}'(s) \in (0,1]$, note that we have $\mathcal{E}_i(S_{ik}-) \leq \mathcal{E}_i(S_{ik-1}) + T_{ik},\ k = 1,2,\ldots,$ implying that the ith subject's effective age just before the kth event occurrence, represented by $\mathcal{E}_i(S_{ik}-)$, is at most the subject's effective age just after the $(k-1)$th event occurrence, which is $\mathcal{E}_i(S_{ik-1})$, plus the inter-occurrence time between the $(k-1)$th and the kth events. This means that the effect of the performed intervention after an event occurrence is to make the subject 'age' at a slower rate relative to the elapsed calendar time.

The generality and scope of this class of models has been established in Peña and Hollander (2004) for the case where $Z_i = 1,\ i = 1,2,\ldots,n$. With the added feature of having the frailty component, this class of models subsumes many existing models in the literature. Below we describe some existing models which are special cases of the general class of models. For more details, see Peña and Hollander (2004). It suffices to describe these models by setting $n = 1$.

Example 2.1: By letting $Z = 1$ (no frailty), $\rho(k;\alpha) = 1,\ k \in \mathbf{Z}_+,\ \psi(w) = 1$, and $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$, we obtain the model where the inter-occurrence times $T_j,\ j = 1,2,\ldots$, are i.i.d. with common hazard rate function $\lambda_0(\cdot)$. This is one of the models examined in Gill (1981) and Peña, et al. (2001). By allowing the frailty Z to have a non-degenerate distribution, one obtains a model that allows for associations among the inter-occurrence times of the subject, a model also considered in Peña, et al. (2001) and Wang and Chang (1999). ‖

Example 2.2: The extended Cox proportional hazards model considered by Prentice, Williams and Peterson (1981), Lawless (1987), and Aalen and Husebye (1991) is a special case of the general model obtained by setting $Z = 1,\ \rho(k;\alpha) = 1,\ k \in \mathbf{Z}_+,\ \mathcal{E}(s) = s - S_{N^\dagger(s-)}$, and $\psi(w) = \exp(w)$. ‖

Example 2.3: Still with $Z = 1,\ \rho(k) = 1,\ k \in \mathbf{Z}$, and $\mathcal{E}(s) = s$, a model examined by Prentice, et al. (1981), Brown and Proschan (1983), and Lawless (1987) arises from the general class of models. This resulting model is referred to in the reliability literature as a minimal repair model, since it arises by 'restoring a system to the state just before it failed (minimally repaired)' whenever the system fails. $\|$

Example 2.4: The Gail, Santner and Brown (1980) Markovian model for tumor occurrence becomes a special case of model (5) by taking $Z = 1$, $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$ and $\rho(k;\alpha) = \alpha - k + 1$, which also coincides with the Jelinski and Moranda (1972) software reliability model, but with the added feature that covariate effects have been incorporated. $\|$

Example 2.5: Let I_1, I_2, $I_3, \ldots$ be a sequence of i.i.d. Bernoulli random variables with success probability p. Define the process $\{\eta(s) : s \in [0,\tau]\}$ via $\eta(s) = \sum_{i=1}^{N^\dagger(s)} I_i$. Also, let $0 \equiv \Gamma_0 < \Gamma_1 < \Gamma_2 < \ldots$ be defined according to $\Gamma_k = \min\{j > \Gamma_{k-1} : I_j = 1\},\ k = 1,2,3,\ldots$. By setting $Z = 1$, $\rho(k;\alpha) = 1$ and $\mathcal{E}(s) = s - S_{\Gamma_{\eta(s-)}}$, we obtain

$$\lambda(s|Z,\mathbf{X}) = \lambda_0\left(s - S_{\Gamma_{\eta(s-)}}\right)\psi(\beta^{\mathrm{t}}\mathbf{X}(s)). \tag{6}$$

This is the Brown and Proschan (1983) minimal repair model in reliability. This is the model studied by Whitaker and Samaniego (1989) where they pointed out that in estimating the reliability function, it suffices to know the inter-failure times *and* the repair modes to achieve model identifiability. If the success probability p is made to depend on the time of event occurrence, the Block, Borges and Savits (1985) model obtains (see also Hollander, Presnell and Sethuraman (1992) and Presnell, Hollander and Sethuraman (1994)). Note in this example that the Γ_ks represent event occurrences in which intervention causes the unit to acquire an effective age of zero. Furthermore, $S_{\Gamma_{\eta(s-)}}$ is the last time prior to s that the subject had an effective age of zero. More generally, the class of models also subsumes the general repair model of Dorado, Hollander, and Sethuraman (1997), which in turns include as special cases models of Kijima (1989), Last and Szekli (1998), Baxter, Kijima, and Tortorella (1996), and Stadje and Zuckerman (1991). For more discussion on this, see the more detailed technical report Peña, Slate, and Gonzalez (2003). $\|$

Example 2.6: An example where $\rho(\cdot;\alpha)$ is not identically unity is provided by taking $\rho[N^\dagger(s-);\alpha] = \alpha^{N^\dagger(s-)}$ for some $\alpha \in \Re_+$. If, additionally, we take $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$, the resulting model postulates that the effect of accumulating event occurrences is a proportional increase (if $\alpha > 1$) in the intensity rate relative to the preceding intensity rate. This could serve as a simple and natural model for the weakening of the subject caused by the accumulating number of event occurrences. Under this specification, and assuming the exponential form for ψ, the intensity process becomes $\lambda(s|\mathbf{X}) = \lambda_0(s - S_{N^\dagger(s-)})\alpha^{N^\dagger(s-)} \exp\{\beta^{\mathrm{t}}\mathbf{X}(s)\}$. Clearly, the above specification could be coupled to the other forms of $\mathcal{E}(s)$ considered in the preceding examples. $\|$

Example 2.7: Another generalization obtains via $\rho[N^\dagger(s-);\alpha] = \max\{\alpha - g[N^\dagger(s-)], 0\}$, where α is some positive real number, and $g(\cdot)$ is some nondecreasing function. One could interpret the parameter α as an initial measure of the unit's susceptibility to events, and $g(\cdot)$ specifies the rate at which this unit is becoming stronger as the event occurrences accumulate. If we take $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$, the resulting model possesses the interesting property that the unit's defects contribute to the event occurrence intensity multiplicatively through the baseline hazard rate function $\lambda_0(\cdot)$. If $g[N^\dagger(s-)] = N^\dagger(s-)$ and $\lambda_0(s) = \lambda_0$, where λ_0 is some positive constant, then the Gail, et al. (1980) tumor occurrence model and the Jelinski and Moranda (1972) software reliability model are obtained. $\|$

Example 2.8: Load-sharing models occur in a variety of situations dealing with coherent systems, computer networks, materials science such as fibrous composites, etc. A popular load-sharing model is the equal load-share model considered recently in Kvam and Peña (2003). A setting in which this occurs is when one considers a K-component parallel system consisting of identical components, and the event of interest for this system is the occurrence of a component failure. Failed components are not replaced, and when a component fails, the load of the system is redistributed equally over the remaining functioning components. To model this, we let $\alpha = (\alpha_0 \equiv 1, \alpha_1, \ldots, \alpha_{K-1})$ be an unknown vector of constants, and we model the hazard rate of event occurrence at calendar time s via $\lambda(s) = \lambda_0(s)[K - N^\dagger(s-)]\alpha_{N^\dagger(s-)}$, where $\lambda_0(\cdot)$ is the hazard rate of each component at time zero and $N^\dagger(s)$ denotes the number of components that have failed up to time s. This model is then a special case of the general model with $\mathcal{E}(s) = s$, $\rho(j;\alpha) = (K-j)\alpha_j$, and one has the added flexibility of also incorporating a link function involving covariates if such are observed, as well as frailty components which could model unobserved operating environment factors. Statistical inference issues for this model without covariates and frailties, such as estimating α and the associated baseline survivor function of $\lambda_0(\cdot)$, were considered in Kvam and Peña (2003).

$\|$

3 Estimation of Parameters: Model without Frailties

In this section we address the problem of estimating the model parameters $\Lambda_0(\cdot) = \int_0^{\cdot} \lambda_0(w)\mathrm{d}w$, α and β for the model where it is assumed that $Z_i \equiv 1$, that is, the model without frailties. Thus, the model of interest has intensity process

$$\lambda_i(s|\mathbf{X}_i) = \lambda_0[\mathcal{E}_i(s)]\, \rho[N_i^\dagger(s-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(s)). \tag{7}$$

The observables for the n subjects, which now include the observable effective age processes, are $\left\{\left(\mathbf{X}_i(s), N_i^\dagger(s), Y_i^\dagger(s), \mathcal{E}_i(s)\right) : 0 \le s \le s^*\right\}$, $i = 1, 2, \ldots, n$, where $N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \le s, S_{ij} \le \tau_i\}$ and $Y_i^\dagger(s) = I\{\tau_i \ge s\}$. The statistical identifiability of this class of models without frailties has been established in Theorem 1 of Peña and Hollander (2004). The two basic conditions to achieve identifiability, aside from the non-triviality of $\psi(\cdot)$ and sufficient variability on $\mathbf{X}$, are that for each value of the parameter set $(\lambda_0(\cdot), \alpha, \beta)$, the support of $\mathcal{E}(S_1)$ should contain $[0, \tau]$, and that $\rho(\cdot, \cdot)$ should satisfy the condition that $\rho(k; \alpha^{(1)}) = \rho(k; \alpha^{(2)})$ for each $k \in \{0, 1, 2, \ldots\}$ implies $\alpha^{(1)} = \alpha^{(2)}$. These two conditions are henceforth assumed to hold.

For this model, letting $A_i^\dagger(s) = \int_0^s Y_i^\dagger(v)\, \lambda_0[\mathcal{E}_i(v)]\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \mathrm{d}v$, then with respect to the filtration $\mathbf{F}$, the vector of processes $\mathbf{M}^\dagger = (M_1^\dagger, \ldots, M_n^\dagger) = \mathbf{N}^\dagger - \mathbf{A}^\dagger = (N_1^\dagger - A_1^\dagger, \ldots, N_n^\dagger - A_n^\dagger)$ consists of orthogonal square-integrable martingales with predictable quadratic covariation processes $\langle M_{i_1}^\dagger, M_{i_2}^\dagger\rangle(s) = A_{i_1}^\dagger(s)\, I\{i_1 = i_2\}$. The usual martingale theory developed by Aalen (1978), Gill (1980), Andersen and Gill (1982), and others (cf., Fleming and Harrington (1991) and Andersen, et al. (1993)) does not apply directly for the purpose of estimating $\Lambda_0(\cdot)$. The reason is that the $\lambda_0(\cdot)$ appearing in $A_i^\dagger(\cdot)$ is time-transformed by the observable predictable process $\mathcal{E}_i(\cdot)$, while of interest is to estimate $\Lambda_0(t)$ for a given t. It is tempting and would seem natural to simply define new processes involving the gap times between the event occurrences. However, as pointed out in Peña, et al. (2001), this approach does not work since the resulting processes no longer satisfy martingale properties owing to the effect of the sum-quota accrual scheme.

The technique utilized in Peña, et al. (2001), extending an idea of Sellke (1988) and Gill

(1981), is to define a doubly-indexed process $Z_i(s,t) = I\{\mathcal{E}_i(s) \leq t\}$, $i = 1, 2, \ldots, n$. The index s represents calendar time, which is the natural time of data accrual; while the index t represents gap times. This process indicates whether at calendar time s, the effective age of the ith subject is no more than t. For $i = 1, 2, \ldots, n$, define also the doubly-indexed processes

$$N_i(s,t) = \int_0^s Z_i(v,t) N_i^\dagger(\mathrm{d}v) \quad \text{and} \quad A_i(s,t) = \int_0^s Z_i(v,t) A_i^\dagger(\mathrm{d}v);$$

$$M_i(s,t) = N_i(s,t) - A_i(s,t) = \int_0^s Z_i(v,t) M_i^\dagger(\mathrm{d}v).$$

For a given t, by utilizing the martingale property of $M_i^\dagger$ and the predictability of $Z_i(\cdot,t)$, the process $M_i(\cdot,t)$ is a square-integrable zero-mean martingale; however, for fixed s, the process $M_i(s,\cdot)$ is *not* a martingale, but nevertheless, it also has mean zero.

A critical result is an equivalent expression for $A_i(s,t)$ which involves $\lambda_0(t)$ directly, instead of its time-transformed version. To reveal this expression, define for $j = 1, 2, \ldots, K_i + 1$ the processes

$$\mathcal{E}_{ij-1}(v) = \mathcal{E}_i(v) I\{S_{ij-1} < v \leq S_{ij}\} \quad \text{on} \quad \{Y_i^\dagger(v) > 0\}. \tag{8}$$

Thus, $\mathcal{E}_{ij-1}(\cdot)$ is the restriction of $\mathcal{E}_i(\cdot)$ on the jth interval bounded by successive event occurrence times for the ith subject. Note that on $(S_{ij-1}, S_{ij}]$, the paths of $\mathcal{E}_{ij-1}(\cdot)$ are one-to-one, so its inverse exists; and furthermore, it is also differentiable. We now provide the alternative expression for $A_i(s,t)$. The proof of this result is analogous to that in Peña, Strawderman and Hollander (2000); see also Stocker and Peña (2003). To make our notation more concise, with $\mathcal{E}'_{ij}(s) = \frac{d}{ds}\mathcal{E}_{ij}(s)$, we define

$$\varphi_{ij}(s;\alpha,\beta) \equiv \frac{\rho[N_i^\dagger(s-);\alpha]\,\psi[\beta^{\mathrm{t}}\mathbf{X}_i(s)]}{\mathcal{E}'_{ij}(s)}. \tag{9}$$

Proposition 1 *For each $i = 1, 2, \ldots, n$, $A_i(s,t) = \int_0^t Y_i(s,w)\lambda_0(w)\mathrm{d}w$, where*

$$Y_i(s,w) \equiv Y_i(s,w|\alpha,\beta) = \sum_{j=1}^{N_i^\dagger(s-)} I_{(\mathcal{E}_{ij-1}(S_{ij-1}),\ \mathcal{E}_{ij-1}(S_{ij})]}(w)\,\varphi_{ij-1}\left(\mathcal{E}_{ij-1}^{-1}(w);\alpha,\beta\right) +$$

$$I_{(\mathcal{E}_{iN_i^\dagger(s-)}(S_{iN_i^\dagger(s-)}),\ \mathcal{E}_{iN_i^\dagger(s-)}(s\wedge\tau_i)]}(w)\,\varphi_{iN_i^\dagger(s-)}\left(\mathcal{E}_{iN_i^\dagger(s-)}^{-1}(w);\alpha,\beta\right).$$

Using Proposition 1, we have the identity $M_i(s,t) = N_i(s,t) - \int_0^t Y_i(s,w)\Lambda_0(\mathrm{d}w)$, $i = 1, 2, \ldots, n$, so that $\sum_{i=1}^n M_i(s, \mathrm{d}w) = \sum_{i=1}^n N_i(s, \mathrm{d}w) - S_0(s,w)\Lambda_0(\mathrm{d}w)$, where

$$S_0(s,t) \equiv S_0(s,t|\alpha,\beta) = \sum_{i=1}^n Y_i(s,t|\alpha,\beta). \tag{10}$$

Because $\sum_{i=1}^n M_i(s,\mathrm{d}w)$ has mean zero, a method-of-moments 'estimator' of $\Lambda_0(t)$, given (α,β) is

$$\hat{\Lambda}_0(s,t;\alpha,\beta) = \int_0^t \left\{ \frac{J(s,w|\alpha,\beta)}{S_0(s,w|\alpha,\beta)} \right\} \left\{ \sum_{i=1}^n N_i(s,\mathrm{d}w) \right\}, \tag{11}$$

with $J(s,w|\alpha,\beta) = I\{S_0(s,w|\alpha,\beta) > 0\}$ and with the convention that $0/0 = 0$. Notice that this 'estimator' is of the same flavor as the Nelson-Aalen estimator or the Aalen-Breslow estimator in single-event settings, although it should be pointed out that the derivation as well as the structure of the processes are quite different.

Next we develop the profile likelihood for (α,β) from which the estimator of (α,β) will be obtained. Following Jacod (1975) (see also Andersen, et al., 1993), if the distribution G of τ does not involve the model parameters, then the likelihood process associated with the observables for the general model without frailties is

$$\begin{aligned} L^\dagger(s|\lambda_0(\cdot),\alpha,\beta) = &\left\{ \prod_{i=1}^n \prod_{v=0}^s \left[Y_i^\dagger(v)\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \lambda_0[\mathcal{E}_i(v)] \right]^{N_i^\dagger(\Delta v)} \right\} \times \\ &\left\{ \exp\left[-\sum_{i=1}^n \int_0^s Y_i^\dagger(v)\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \lambda_0[\mathcal{E}_i(v)]\, \mathrm{d}v \right] \right\}. \end{aligned} \tag{12}$$

The argument of the exponential function could be re-expressed via

$$\sum_{i=1}^n \int_0^s Y_i^\dagger(v)\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \lambda_0[\mathcal{E}_i(v)]\mathrm{d}v = \sum_{i=1}^n A_i(s,\infty) = \int_0^\infty S_0(s,w|\alpha,\beta)\, \Lambda_0(\mathrm{d}w).$$

Since from (11), we have $\hat{\Lambda}_0(s,\mathrm{d}w|\alpha,\beta) = \sum_{i=1}^n N_i(s,\mathrm{d}w)/S_0(s,w|\alpha,\beta)$, it therefore follows that $\int_0^\infty S_0(s,w|\alpha,\beta)\hat{\Lambda}_0(s,\mathrm{d}w|\alpha,\beta) = \sum_{i=1}^n N_i(s,\infty)$, which is independent of (α,β). Upon substituting the 'estimator' $\hat{\Lambda}_0(s,t|\alpha,\beta)$ for $\Lambda_0(t)$ in the argument of the exponential function in (12), the resulting term will not contribute to the profile likelihood for (α,β).

On the other hand, substituting $\hat{\Lambda}_0(s,w|\alpha,\beta)$ for $\Lambda_0(w)$ in the first term of (12), we obtain the relevant portion of the profile likelihood of (α,β) to be

$$L_p(s|\alpha,\beta) = \prod_{i=1}^n \prod_{j=1}^{N_i^\dagger(s)} \left[\frac{\rho(j-1;\alpha)\, \psi[\beta^{\mathrm{t}}\mathbf{X}_i(S_{ij})]}{S_0[s,\mathcal{E}_i(S_{ij})|\alpha,\beta]} \right]^{\Delta N_i^\dagger(S_{ij})}. \tag{13}$$

This process could also be viewed as the partial likelihood process for (α, β), which is a generalization of the partial likelihood for the Cox model (cf., Cox (1972; 1975); Andersen and Gill (1982)). The logarithm of the profile likelihood could be conveniently expressed in integral form via

$$l_P(s|\alpha, \beta) = \sum_{i=1}^{n} \int_0^s \left[\log \rho[N_i^\dagger(v-); \alpha] + \log \psi(\beta^{\mathrm{t}} \mathbf{X}_i(v)) - \log S_0(s, \mathcal{E}_i(v)|\alpha, \beta)\right] N_i^\dagger(\mathrm{d}v). \quad (14)$$

From this profile likelihood, the estimators of α and β will be obtained. It is easy to see that the estimating equations for the profile maximum likelihood estimators are

$$\sum_{i=1}^{n} \int_0^{s^*} \left[\frac{\frac{\partial}{\partial \alpha} \rho[N_i^\dagger(v-); \alpha]}{\rho[N_i^\dagger(v-); \alpha]} - \frac{\frac{\partial}{\partial \alpha} S_0(s, \mathcal{E}_i(v)|\alpha, \beta)}{S_0(s, \mathcal{E}_i(v)|\alpha, \beta)}\right] N_i^\dagger(\mathrm{d}v) = \mathbf{0}; \quad (15)$$

$$\sum_{i=1}^{n} \int_0^{s^*} \left[\frac{\frac{\partial}{\partial \beta} \psi(\beta^{\mathrm{t}} \mathbf{X}_i(v))}{\psi(\beta^{\mathrm{t}} \mathbf{X}_i(v))} - \frac{\frac{\partial}{\partial \beta} S_0(s, \mathcal{E}_i(v)|\alpha, \beta)}{S_0(s, \mathcal{E}_i(v)|\alpha, \beta)}\right] N_i^\dagger(\mathrm{d}v) = \mathbf{0}. \quad (16)$$

Because $N_i^\dagger(\cdot)$ is a step process with a finite number of jumps, then both of these estimating equations are finite sums with respect to the calendar times S_{ij}s. Also, just like estimating equations in simpler models, such as for the Cox proportional hazards model, it is clear that numerical techniques will be needed to obtain the estimates $\hat{\alpha}$ and $\hat{\beta}$.

The structure of the estimating equations in (15) and (16) can be better understood by introducing new notation. For $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, N_i^\dagger(s)$, and recalling the definition of the function $\varphi_{ij}(\cdot; \alpha, \beta)$ in (9), define

$$Q_{ij}(s, w|\alpha, \beta) = I_{(\mathcal{E}_{ij-1}(S_{ij-1}),\ \mathcal{E}_{ij-1}(S_{ij})]}(w)\, \varphi_{ij-1}\left(\mathcal{E}_{ij-1}^{-1}(w); \alpha, \beta\right); \quad (17)$$

$$R_i(s, w|\alpha, \beta) = I_{(\mathcal{E}_{iN_i^\dagger(s-)}(S_{iN_i^\dagger(s-)}),\ \mathcal{E}_{iN_i^\dagger(s-)}(\min(s, \tau_i))]}(w)\, \varphi_{iN_i^\dagger(s-)}\left(\mathcal{E}_{iN_i^\dagger(s-)}^{-1}(w); \alpha, \beta\right). \quad (18)$$

Note that these processes satisfy the technical and crucial condition of predictability. Using these processes, $S_0(s, w|\alpha, \beta)$ could be re-expressed via

$$S_0(s, w|\alpha, \beta) = \sum_{i=1}^{n} \left\{\sum_{j=1}^{N_i^\dagger(s-)} Q_{ij}(s, w|\alpha, \beta) + R_i(s, w|\alpha, \beta)\right\}. \quad (19)$$

Observe that the Q_{ij}s can be interpreted as the contributions of the uncensored values, while the R_is are the contributions of the right-censored values. With a slight change in notation above, we therefore observe that this mirrors the single-event situation.

For notation, let $\rho^{(\alpha)}(\cdot;\alpha) = \partial\rho(\cdot;\alpha)/\partial\alpha$, $\psi'(\cdot)$ be the derivative of $\psi(\cdot)$, and

$$\mathbf{V}(j;\alpha) = \frac{\rho^{(\alpha)}(j;\alpha)}{\rho(j;\alpha)} \quad \text{and} \quad \mathbf{W}(\mathbf{x};\beta) = \frac{\mathbf{x}\psi'(\beta^{\mathrm{t}}\mathbf{x})}{\psi(\beta^{\mathrm{t}}\mathbf{x})}.$$

Then taking the partial derivatives (more appropriately, the gradients) of $S_0(s,w|\alpha,\beta)$ with respect to α and β, we obtain

$$\begin{aligned}
\frac{\partial}{\partial\alpha}S_0(s,w|\alpha,\beta) &= \sum_{i=1}^{n}\left\{\sum_{j=1}^{N_i^\dagger(s-)}\mathbf{V}(j-1;\alpha)\,Q_{ij}(s,w|\alpha,\beta) + \mathbf{V}(N_i^\dagger(s-);\alpha)\,R_i(s,w|\alpha,\beta)\right\};\\
\frac{\partial}{\partial\beta}S_0(s,w|\alpha,\beta) &= \sum_{i=1}^{n}\left\{\sum_{j=1}^{N_i^\dagger(s-)}\mathbf{W}_{ij}(w;\beta)\,Q_{ij}(s,w|\alpha,\beta) + \mathbf{W}_{iN_i^\dagger(s-)}(w;\beta)\,R_i(s,w|\alpha,\beta)\right\},
\end{aligned}$$

where for brevity, $\mathbf{W}_{ij}(w;\beta) = \mathbf{W}(\mathbf{X}_i(\mathcal{E}_{ij-1}^{-1}(w));\beta)$. These expressions simplify when specific forms of $\rho(\cdot;\alpha)$ and $\psi(\cdot)$ are taken, or if the covariate process are time-independent. One possible choice is $\rho(j;\alpha) = \alpha^j$, leading to $\mathbf{V}(j;\alpha) = j/\alpha$; and for $\psi(\cdot)$ a common choice is $\psi(w) = \exp(w)$, for which $\psi'(w) = \psi(w)$, so we obtain $\mathbf{W}(\mathbf{X}_i(\mathcal{E}_{ij-1}^{-1}(w));\beta) = \mathbf{X}_i(\mathcal{E}_{ij-1}^{-1}(w))$, and, if this is coupled with the assumption that the covariate vector process is time-independent, then $\mathbf{X}_i(\mathcal{E}_{ij-1}^{-1}(w)) = \mathbf{X}_i$.

To obtain further simplification for the moment, let us assume that $\mathbf{X}_i$ are time-independent, $\rho(j;\alpha) = \alpha^j$, and $\psi(w) = \exp(w)$. Under these assumptions, letting

$$A(s,w|\alpha,\beta) = \frac{1}{\alpha}\frac{\sum_{i=1}^{n}\left\{\sum_{j=1}^{N_i^\dagger(s-)}(j-1)\,Q_{ij}(s,w|\alpha,\beta) + N_i^\dagger(s-)\,R_i(s,w|\alpha,\beta)\right\}}{\sum_{i=1}^{n}\left\{\sum_{j=1}^{N_i^\dagger(s-)}Q_{ij}(s,w|\alpha,\beta) + R_i(s,w|\alpha,\beta)\right\}};$$

$$\mathbf{B}(s,w|\alpha,\beta) = \frac{\sum_{i=1}^{n}\mathbf{X}_i\left\{\sum_{j=1}^{N_i^\dagger(s-)}Q_{ij}(s,w|\alpha,\beta) + R_i(s,w|\alpha,\beta)\right\}}{\sum_{i=1}^{n}\left\{\sum_{j=1}^{N_i^\dagger(s-)}Q_{ij}(s,w|\alpha,\beta) + R_i(s,w|\alpha,\beta)\right\}},$$

it is easy to see that the estimating equations in (15) and (16) become

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{N_i^\dagger(s^*-)}\left[\frac{j-1}{\alpha} - A(s^*,\mathcal{E}_{ij-1}(S_{ij})|\alpha,\beta)\right]\Delta N_i^\dagger(S_{ij}) &= 0;\\
\sum_{i=1}^{n}\sum_{j=1}^{N_i^\dagger(s^*-)}\left[\mathbf{X}_i - \mathbf{B}(s^*,\mathcal{E}_{ij-1}(S_{ij})|\alpha,\beta)\right]\Delta N_i^\dagger(S_{ij}) &= \mathbf{0}.
\end{aligned}$$

Upon obtaining the estimators $\hat\alpha$ and $\hat\beta$ as described in the preceding discussion, the estimator of $\Lambda_0(t)$ based on the realizations of the observables over $[0,s^*]$ is obtained by substituting $(\hat\alpha,\hat\beta)$

for (α, β) in the expression of $\hat{\Lambda}_0(s^*, t|\alpha, \beta)$ given in (11). Thus,

$$\hat{\Lambda}_0(s^*, t) = \int_0^t \left\{ \frac{J(s^*, w|\hat{\alpha}, \hat{\beta})}{S_0(s^*, w|\hat{\alpha}, \hat{\beta})} \right\} \left\{ \sum_{i=1}^n N_i(s^*, \mathrm{d}w) \right\}. \tag{20}$$

Finally, for an estimator of the baseline survivor function associated with $\Lambda_0(\cdot)$ defined via $\bar{F}_0(t) = \exp\{-\Lambda_0(t)\}$, by the product-integral representation and the substitution principle, we obtain

$$\hat{\bar{F}}_0(s^*, t) = \prod_{w=0}^{t} \left[1 - \hat{\Lambda}_0(s^*, \mathrm{d}w)\right] = \prod_{w=0}^{t} \left[1 - \frac{\sum_{i=1}^n N_i(s^*, \mathrm{d}w)}{S_0(s^*, w|\hat{\alpha}, \hat{\beta})}\right]. \tag{21}$$

This estimator is of a product-limit type analogous to those arising in the estimation of the baseline survivor function in the Cox proportional hazards model or the multiplicative intensity model (Cox 1972; Andersen and Gill 1982).

For the i.i.d. interoccurrence times model in Example 2.1, which obtains when $\psi(w) = 1$ (no covariate effects), $\rho(w) = 1$ (no effects of accumulating event occurrences), and $\mathcal{E}_i(s) = s - S_{N_i^\dagger(s-)}$ (upon each event occurrence, effective age is reset to zero, so this is just the backward recurrence time), the estimator of $\bar{F}_0(t)$ in (21) simplifies to that considered in Peña, et al. (2001). Note, in particular, that for this special model, $\mathcal{E}_i'(s) = 1$, and since $\mathcal{E}_{ij-1}(S_{ij}) = S_{ij} - S_{ij-1} = T_{ij}$, then the process $Y_i(s, w)$ simplifies to $Y_i(s, w) = \sum_{j=1}^{N_i^\dagger(s-)} I\{T_{ij} \geq w\} + I\{\min(s, \tau_i) - S_{iN_i^\dagger(s-)} \geq w\}$, which is the natural at-risk process for the gap times over the observation period $[0, s]$.

4 Estimation of Parameters: Model with Frailties

In this section we consider the estimation of the parameters when the class of models includes frailties. It will be assumed that the frailties $Z_1, Z_2, \ldots, Z_n$ are i.i.d. from a distribution $H(\cdot|\xi)$ where $\xi \in \Xi \subseteq \Re^r$. A common choice for this H, which we adopt here, is the gamma distribution with unit mean and variance $1/\xi$, $H = \text{Gamma}(\xi, \xi)$. Imposing the restriction that the gamma shape and scale parameters are identical is needed to have model identifiability. Recall at this stage that the Z_is are not observed. For the model at hand, the conditional intensity function is as given in (5), which, for convenience, is again displayed below:

$$\lambda_i(s|Z_i, \mathbf{X}_i) = Z_i\, \lambda_0[\mathcal{E}_i(s)]\, \rho[N_i^\dagger(s-); \alpha]\, \psi(\beta^{\mathrm{t}} \mathbf{X}_i(s)).$$

In estimating the model parameters ξ, $\Lambda_0(\cdot)$, α, and β, we generalize and extend the approach implemented in Peña, et al. (2001) which dealt with the frailty model without covariates, and without the $\rho(\cdot;\alpha)$ term, and with $\mathcal{E}_i(s) = s - S_{iN_i^\dagger(s-)}$. The computations of the estimates will be facilitated through the expectation-maximization (EM) algorithm introduced by Dempster, Laird, and Rubin (1977), and implemented in counting process frailty models by Nielsen, et al. (1992). The main ingredients of this algorithm for the general class of recurrent event models are as follows: Given $(\Lambda_0(\cdot), \alpha, \beta)$ and $\mathbf{D}(s^*) \equiv (\mathbf{D}_1(s^*), \ldots, \mathbf{D}_n(s^*))$, the conditional expectation of Z_i is

$$\mathbf{E}\{Z_i|\Lambda_0(\cdot), \alpha, \beta\} = \frac{\xi + N_i^\dagger(s^*)}{\xi + \int_0^{s^*} Y_i^\dagger(v)\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \lambda_0[\mathcal{E}_i(v)]\, \mathrm{d}v}. \tag{22}$$

Furthermore, following the development of the 'estimator' $\hat{\Lambda}_0(s^*, \cdot|\alpha, \beta)$ for the model without frailties in the preceding subsection, given $\mathbf{Z} = (Z_1, \ldots, Z_n)$, α, β, and the data $\mathbf{D}$, the 'estimator' of $\Lambda_0(\cdot)$ is given by

$$\hat{\Lambda}_0(s^*, t|\mathbf{Z}, \alpha, \beta) = \int_0^t \left\{ \frac{J(s^*, w|\mathbf{Z}, \alpha, \beta)}{S_0(s^*, w|\mathbf{Z}, \alpha, \beta)} \right\} \left\{ \sum_{i=1}^n N_i(s^*, \mathrm{d}w) \right\}, \tag{23}$$

where $J(s, w|\mathbf{Z}, \alpha, \beta) = I\{S_0(s, w|\mathbf{Z}, \alpha, \beta) > 0\}$ with $S_0(s, w|\mathbf{Z}, \alpha, \beta) = \sum_{i=1}^n Z_i Y_i(s, w|\alpha, \beta)$. Analogously to the estimating equations for α and β in the model without frailties in (15) and (16), given $\mathbf{Z}$ and $\hat{\Lambda}_0(s^*, \cdot|\mathbf{Z}, \alpha, \beta)$, we may estimate α and β by solving the estimating equations

$$\sum_{i=1}^n \int_0^{s^*} \left[\frac{\frac{\partial}{\partial\alpha}\rho[N_i^\dagger(v-);\alpha]}{\rho[N_i^\dagger(v-);\alpha]} - \frac{\frac{\partial}{\partial\alpha} S_0(s, \mathcal{E}_i(v)|\mathbf{Z}, \alpha, \beta)}{S_0(s, \mathcal{E}_i(v)|\mathbf{Z}, \alpha, \beta)} \right] N_i^\dagger(\mathrm{d}v) = \mathbf{0}; \tag{24}$$

$$\sum_{i=1}^n \int_0^{s^*} \left[\frac{\frac{\partial}{\partial\beta}\psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))}{\psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))} - \frac{\frac{\partial}{\partial\beta} S_0(s, \mathcal{E}_i(v)|\mathbf{Z}, \alpha, \beta)}{S_0(s, \mathcal{E}_i(v)|\mathbf{Z}, \alpha, \beta)} \right] N_i^\dagger(\mathrm{d}v) = \mathbf{0}. \tag{25}$$

On the other hand, by integrating out $\mathbf{Z}$ according to its joint (gamma) distribution in the joint likelihood function, the marginal profile likelihood for ξ, given $(\Lambda_0(\cdot), \alpha, \beta)$, is obtained as

$$\begin{aligned} L_P(s^*|\xi, \alpha, \beta, \Lambda_0(\cdot)) = \prod_{i=1}^n \Bigg\{ & \left[\frac{\Gamma(\xi + N_i^\dagger(s^*))}{\Gamma(\xi)} \right] \times \\ & \left[\frac{\xi}{\xi + \int_0^{s^*} Y_i^\dagger(v)\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \lambda_0[\mathcal{E}_i(v)]\, \mathrm{d}v} \right]^{\xi + N_i^\dagger(s^*)} \times \\ & \left(\prod_{v=0}^{s^*} \left[\frac{Y_i^\dagger(v)\, \rho[N_i^\dagger(v-);\alpha]\, \psi(\beta^{\mathrm{t}}\mathbf{X}_i(v))\, \lambda_0[\mathcal{E}_i(v)]}{\xi} \right]^{N_i^\dagger(\Delta v)} \right) \Bigg\}. \end{aligned} \tag{26}$$

For a given $(\Lambda_0(\cdot), \alpha, \beta)$, this function could be maximized with respect to ξ using numerical maximization algorithms. With these ingredients at hand, the EM recipe for obtaining the estimates of the model parameters in this general model with frailties is described by the following steps:

Step 0 (Initialization): Specify initial estimates $\hat{\xi}^{(0)}$, $\hat{\alpha}^{(0)}$, and $\hat{\beta}^{(0)}$ of ξ, α, and β, respectively. By setting $\hat{Z}_i^{(0)} = 1, i = 1, 2, \ldots, n$, obtain the initial estimate of $\Lambda_0(\cdot)$ via

$$\hat{\Lambda}_0^{(0)}(s^*, t|\hat{\mathbf{Z}}^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)}) = \int_0^t \left\{ \frac{J(s^*, w|\hat{\mathbf{Z}}^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})}{S_0(s^*, w|\hat{\mathbf{Z}}^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})} \right\} \left\{ \sum_{i=1}^n N_i(s^*, \mathrm{d}w) \right\}$$

which is just the 'estimator' in (11) under the model without frailties.

Step 1 (E-step): Given $(\hat{\xi}^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})$, and $\hat{\Lambda}_0^{(0)}(s^*, \cdot|\hat{\mathbf{Z}}^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})$, obtain the estimated frailty values $\hat{Z}_i^{(1)} = \mathbf{E}\left\{ Z_i|\hat{\Lambda}_0^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)} \right\}, i = 1, 2, \ldots, n$, according to the formula in (22). Denote these estimated values by $\hat{\mathbf{Z}}^{(1)} = (\hat{Z}_1^{(1)}, \ldots, \hat{Z}_n^{(1)})$. By exploiting the property that the estimator $\hat{\Lambda}_0(\cdot)$ is a step function, these quantities could be obtained according to the following expressions: For $i = 1, 2, \ldots, n$,

$$\hat{Z}_i^{(1)} = \frac{\hat{\xi}^{(0)} + N_i^\dagger(s^*)}{\hat{\xi}^{(0)} + \hat{A}_i(s^*; \hat{\Lambda}_0^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})} \tag{27}$$

where, with $t_{(1)} < t_{(2)} < \ldots < t_{(D)}$ being the D distinct jump times of $\hat{\Lambda}_0^{(0)}(s^*, \cdot)$ and $\hat{\lambda}_0^{(0)}(s^*, t_{(l)}) = \hat{\Lambda}_0^{(0)}(s^*, t_{(l)}) - \hat{\Lambda}_0^{(0)}(s^*, t_{(l)}-)$ is the jump of $\hat{\Lambda}_0^{(0)}(s^*, \cdot)$ at $t = t_{(l)}$, we have

$$\hat{A}_i^{(0)} \equiv \hat{A}_i(s^*; \hat{\Lambda}_0^{(0)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)}) = \textstyle\sum_{l=1}^D Y_i(s^*, t_{(l)}|\hat{\alpha}^{(0)}, \hat{\beta}^{(0)})\ \hat{\lambda}_0^{(0)}(s^*, t_{(l)}). \tag{28}$$

Step 2 (M-step #1): Applying formula (23), obtain $\hat{\Lambda}_0^{(1)}(s^*, t|\hat{\mathbf{Z}}^{(1)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})$.

Step 3 (M-step #2): After substituting $\hat{\mathbf{Z}}^{(1)}$ for $\mathbf{Z}$ in the estimating equations (24) and (25), obtain the solutions of these equations and denote them by $\hat{\alpha}^{(1)}$ and $\hat{\beta}^{(1)}$.

Step 4 (M-step #3): Replacing $\Lambda_0(\cdot)$ in (26) by $\hat{\Lambda}_0^{(1)}(s^*, \cdot|\hat{\mathbf{Z}}^{(1)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})$ from **Step 2**, and (α, β) by $(\hat{\alpha}^{(1)}, \hat{\beta}^{(1)})$, maximize the resulting (estimated) marginal likelihood with respect to ξ to obtain $\hat{\xi}^{(1)}$. The function to be maximized with respect to ξ is given by

$$\prod_{i=1}^n \left\{ \left[\frac{\Gamma(\xi + N_i^\dagger(s^*))}{\Gamma(\xi)} \right] \left[\frac{1}{\xi + \hat{A}_i^{(1)}} \right]^{\xi + N_i^\dagger(s^*)} \xi^\xi \exp\{\hat{B}_i^{(1)}\} \right\} \tag{29}$$

where $\hat{A}_i^{(1)}$ is computed using the formula in (28) and $\hat{B}_i^{(1)}$ is computed via

$$\hat{B}_i^{(1)} = \sum_{j=1}^{N_i^\dagger(s^*)} \log\left\{\rho(j-1;\hat{\alpha}^{(1)})\,\psi(\hat{\beta}^{(1)\mathrm{t}}\mathbf{X}_i(S_{ij}))\,\hat{\lambda}_0^{(1)}(s^*,\mathcal{E}_i(S_{ij})|\hat{\mathbf{Z}}^{(1)},\hat{\alpha}^{(1)},\hat{\beta}^{(1)})\right\}. \tag{30}$$

The maximization of (29) with respect to ξ may be aided by a reparameterization to, for example, $\log(\xi)$, since this will alleviate the problem of negative values when using iterative gradient-based algorithms.

Step 5 (Convergence): Compare the values $(\hat{\xi}^{(1)}, \hat{\mathbf{Z}}^{(1)}, \hat{\Lambda}_0^{(1)}(s^*,\cdot|\hat{\mathbf{Z}}^{(1)},\alpha^{(0)},\beta^{(0)}),\alpha^{(1)},\beta^{(1)})$ with the values $(\hat{\xi}^{(0)}, \hat{\mathbf{Z}}^{(0)}, \hat{\Lambda}_0^{(0)}(s^*,\cdot|\hat{\mathbf{Z}}^{(0)},\alpha^{(0)},\beta^{(0)}),\alpha^{(0)},\beta^{(0)})$ according to some distance function, e.g., Euclidean distance. If the distance between the old and the new values satisfy a certain tolerance criterion, the algorithm terminates and the estimates are the final values in the iteration. If the distance criterion is not satisfied, then replace the old values by the new values, and proceed to **Step 1** of the algorithm. Because of the possibility of very large, possibly infinite, estimates of ξ, corresponding to the situation of approximate 'uncorrelatedness,' when comparing old and new iterates for ξ, we compare instead the associated values for $\eta = \xi/(1+\xi)$ since this ratio takes values in $(0,1]$.

Having obtained an estimator of the baseline hazard function $\Lambda_0(\cdot)$ given by $\hat{\Lambda}_0(s^*,\cdot)$, by virtue of the product integral representation of the survivor curve, the semiparametric estimator of the baseline survivor function $\bar{F}_0(\cdot)$ for this model with frailty is

$$\hat{\bar{F}}_0(s^*,t) = \prod_{\{w:\ w\le t\}}\left[1-\hat{\Lambda}_0(s^*,dw)\right]. \tag{31}$$

A computational implementation of the procedures and algorithms described in Sections 3 and 4 have been implemented in an `R` package (Ihaka and Gentleman 1996) called *gcmrec* in González, Slate, and Peña (2003).

5 Properties of Estimators

5.1 Simulation Design

We performed computer simulation studies to examine numerically the properties of the parameter estimators developed in Sections 3 and 4. The specific goals of these studies are: (i) to examine the effect of sample size (n) on the distributional properties of the estimators; (ii) to examine the bias, variance, and root-mean-square error (rmse) of the estimators; (iii) to examine the performance of the semiparametric estimator of the baseline survivor function $\bar{F}_0$ in terms of its bias function, variance function, and root-mean-squared error function at specified time points. The latter function is based on the loss function $L(\hat{\bar{F}}(t), \bar{F}(t)) = (\hat{\bar{F}}(t) - \bar{F}(t))^2$; (iv) to examine the consequences when data that have been generated with frailty components are analyzed using the model without frailties; and (v) to examine the consequences, such as the loss in efficiency, when data that were generated using the model without frailties are analyzed with methods developed under the model with frailties. For the first three items, simulation runs were performed for both the frailty-less model and for the model with frailty. We describe the settings for the different simulation parameters.

Sample Size: To examine the impact of sample size, we choose three values of n: $n \in \{10, 30, 50\}$. The case of $n = 10$ may not be realistic in biomedical settings, which often have many subjects in the study, but such a small sample size may arise in the reliability and engineering settings, such as for example in the hydraulic data set. Including this small sample size enables us also to examine the limitations of the numerical procedures in obtaining the estimates.

Censoring Mechanism: The censoring variable $\tau_i, i = 1, 2, \ldots, n$, are generated according to a uniform distribution over $[0, B]$ where B is chosen in order that under perfect repair (i.e., $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$) and with $\alpha = 1$, there are, on average, approximately 10 events per unit. Moreover, to place an upper limit to the number of events that could occur for a unit, when the number of events for a unit reaches 50 then we cease observing this unit and set $\tau_i = S_{i,50}$. This has the potential consequence of introducing some bias because this amounts to doing a combination of Type II and random censoring. Nevertheless, because the value of 50 is large enough, we conjecture that the bias introduced is negligible.

Effective Age Function: For the simulation studies we considered an effective age process corresponding to the general minimal repair model (see Example 2.5) with perfect repair probability of 0.6. Since the upper bound for the uniform censoring was determined under the perfect repair model and with $\alpha = 1$ in order to have an average of approximately 10 events per unit, in the simulations the effective average number of events per unit may either be smaller or larger than this prespecified value of 10 owing to the interplay among the baseline hazard rate function (if it is increasing failure rate (IFR) or decreasing failure rate (DFR)), the minimal repairs performed, and the effect of increasing number of event occurrence quantified by α.

Baseline Survivor Function: For the baseline hazard function $\lambda_0(\cdot)$ we choose the flexible and commonly-used Weibull hazard function, with a unit scale parameter and shape parameter (γ) taking values in $\{.9, 2\}$, the former leading to a DFR distribution, and the latter giving rise to an IFR distribution. Note that the estimation procedure proposed is semiparametric, hence the scale and shape parameters of this Weibull baseline distribution are not estimated; on the other hand, see Stocker and Peña (2003) for a parametric treatment of the baseline hazard function.

ρ Function: The ρ function which handles the impact of accumulating event occurrences is assumed to be of form $\rho(k;\alpha) = \alpha^k$ with $\alpha \in \{0.9, 1.0, 1.05\}$, which models the situations where an increasing number of event occurrences has a beneficial effect, has no effect, or has an adverse effect, respectively.

Covariates: We consider a two-dimensional covariate vector (X_1, X_2) with X_1 having a Bernoulli distribution with success probability of 0.5, X_2 having a standard normal distribution, and with X_1 and X_2 stochastically independent. The regression coefficient vector (β_1, β_2) is set to be $(1, -1)$. The fact that the grouping induced by the first covariate is done using a symmetric Bernoulli mechanism lead sometimes to highly asymmetric allocations for some simulation replicates, which was the cause of some convergence problems in the iterative procedure when $n = 10$.

Frailty Component: The parameter ξ of the gamma distribution governing the frailty variable was set to $\{2, 6, \infty\}$, with ∞ corresponding to the absence of frailties. With respect to the parametrization $\xi \mapsto \eta = \xi/(1+\xi)$, these frailty values convert to having $\eta \in \{\frac{2}{3}, \frac{6}{7}, 1\}$.

For each combination of these simulation parameters, 1000 replications were performed. In the analysis, we set $s^* = 10$. Also, to create the bias, variance, and root-mean-squared-error curves for the estimator of the baseline survivor function, we choose the time values that corresponded to the

$[0 : (0.01) : 0.99]$ percentiles of the true baseline distribution function. Associated with the deciles of this true baseline distribution function, we create side-by-side boxplots of the simulated values of $W_n(s^*, t) \equiv \sqrt{n}[\hat{\bar{F}}_0(s^*, t) - \bar{F}_0(t)]$. This enables the graphical and empirical assessment of whether the sampling distributions of these standardized values (at a fixed time point) are converging to a normal distributions. As mentioned earlier, a theoretical treatment of the asymptotic properties of the estimators will be presented in another paper.

5.2 Discussions of Simulation Results

In the discussion of the simulation results that follows, we will focus on the effects of changing n, changing ξ or η, changing α, and changing γ, on the distributional properties of the estimators of α, β, and η, as well as the estimator of the baseline survivor function $\bar{F}_0$. In addition, we also address the consequences of analyzing data that follows the general model with frailty using procedures developed for the general model without frailties, an under-specification; and also consider the impact of over-specification, which is the situation where procedures developed under the model with frailties are utilized to analyze data from a model without frailties. Such analyses will provide information on which type of mis-specification is of a more serious type.

Results of the simulation studies are presented in the following tables. Table 1 summarizes the mean values and standard deviations (i.e., standard errors of the estimates) of the sampling distributions of the estimators of α, β_1, β_2, and η for α values of 0.9, 1.0, and 1.05 as n varies in the set $\{30, 50\}$. We do not anymore show the cases with $n = 10$ to conserve space. Table 2 contains means and standard deviations summaries of the simulation runs pertaining to the under- and over-specified analysis. Table 3 contains plots of the bias and rmse curves for the estimator of $\bar{F}_0$ under the case where $\alpha = .9$ for $\xi \in \{2, 6, \infty\}$ with the plots for different values of n superimposed on each plot frame for two Weibull shape parameter values, $\gamma = 0.9$ and $\gamma = 2.0$. Table 4 contains side-by-side boxplots of $W_n(s^*, t)$ for t-values associated with the deciles of the true survivor function $\bar{F}_0$ for the case where $\alpha = .9$, $\xi = 2$, $\gamma = 0.9$, and as n varies in $\{10, 30, 50\}$. Table 5 contains bias and rmse curves under the mis-specification runs, showing the effect of sample size for different values of α and when $\xi = 2$ and $\gamma = 0.9$.

We note at the outset that one limitation of the computational implementation used in the simulation is that for small sample sizes such as $n = 10$ (not shown in the tables), there were some cases of nonconvergence (indicated by the column NC in the tables), and also the computational procedure in a few runs may not have converged to a maximizing value for the likelihood with respect to the parameter β_1, probably because there were not enough units allocated to one of the groups. For some replicates, the resulting values turned out to be extreme outliers, and the effect of these outliers is evident in the simulated mean values of $\hat{\beta}_1$. When these outliers were removed, the simulated mean values became very close to the true value of β_1.

As is to be expected, for the simulation runs where there was no mis-specification, when the sample size increases, the performance of the estimators of the finite-dimensional parameters, as well as for the baseline survivor function, improved, with the biases decreasing and the standard errors also decreasing. This is also true for the over-specification runs. When the sample size is small, there is considerable over-estimation of $\eta = \xi/(1 + \xi)$, though this bias decreases with increasing sample size. When there is under-specification however, all the estimators are extremely biased (see UVW-runs in Table 2 as well as Table 5, demonstrating the undesirable consequences of committing this under-specification. Regarding the effect of the frailty parameter ξ, for estimating the finite-dimensional parameters, the amount of bias for $n = 30$ and $n = 50$ are negligible. The impact of the ξ is on the standard errors of the estimators, with larger values of ξ translating into less correlation, leading to smaller standard errors for the same sample size. When considering on the other hand the estimator $\hat{\bar{F}}_0$ of the baseline survivor function, by examining the curves in Table 3 we observe that the bias and rmse curves of this estimator decrease as n increases, and the same could also be said as ξ increases. Generally, the bias function is positive, and as is to be expected there is more bias and rmse in the middle portion of the survivor function. Regarding the marginal distributions of $\hat{\bar{F}}_0(t)$, we observe from the side-by-side boxplots in Table 4 that as the sample size increases, the sampling distribution of $W_n(s^*, t)$ in t becomes closer to being symmetric about zero and supports the conjecture that a normal limiting distribution holds. This issue of asymptotic distributions of the estimators will be addressed in a separate paper. In particular, the question of whether the process $\{W_n(s^*, t) : 0 \le t \le t^*\}$ converges weakly to a Gaussian process will be examined.

Some care, however, must be observed when considering the effects of changing α and changing Weibull shape parameter γ in the context of the precision of the estimators because the interplay between these two parameters leads to differing observed number of events. To see this, examine the column $\hat{\mu}_{Ev}$, which represents the mean number of events observed per unit, in Table 1. In this table, we notice that when $\alpha < 1$ and $\gamma < 1$, the latter leading to a DFR Weibull baseline distribution, there tends to be a smaller number of observed events; whereas when $\alpha > 1$ and $\gamma > 1$, the latter making the Weibull baseline IFR, then there tends to be more events observed. These differences in the observed number of events can be explained by taking into account the minimal repair model considered in the simulation. In the first situation for instance, an α value less than unity makes the unit less likely to have events as calendar time increases since more event occurrences become beneficial to the unit and, in addition, when a minimal repair is performed, then the DFR nature (because $\gamma < 1$) of the baseline distribution diminishes the rate of event occurrences thereby lengthening the inter-event times. Because the upper bound B for the uniformly distributed follow-up time τ was determined under $\alpha = 1$ and with a backward recurrence time effective age corresponding to a perfect repair mechanism, the impact of $\alpha < 1$ and $\gamma < 1$ is a smaller number of events compared to the target of approximately 10 events used in deriving B. An analogous argument, but in the opposite direction, holds true when dealing with $\alpha > 1$ and $\gamma > 1$. The impact of the minimal repair effective age and its interplay with a DFR or IFR baseline distribution can be further seen from Table 1 with $\alpha = 1$, where we see that when the baseline distribution is DFR (IFR), the observed number of events per unit is less (more) than the target of approximately 10 events per unit used in deriving B. A fascinating situation is when $\alpha < 1$ and $\gamma > 1$, or when $\alpha > 1$ and $\gamma < 1$, for the effects of α and γ are in opposite directions in the context of event occurrences. Examining the bottom portion of the A-runs in Table 1 and the upper half of the C-runs in Table 1, and with reference to the B-runs in this same table, we observe that for the chosen α and γ values in the simulation, there was a more pronounced effect of the α values compared to the γ

values since when $\alpha = .9$ and $\gamma = 2$, the observed number of events is slightly below 10, whereas when $\alpha = 1.05$ and $\gamma = .9$, the observed number of events is more than 10. The greater effect of α than γ on the mean number of events is not surprising, because γ was partially accommodated in the determination of the upper bound B for the censoring distribution. It is a theoretical challenge, however, to obtain an exact analytical expression for B that will yield a prespecified mean number of event occurrences per unit for a given α, γ, ξ, and specified effective age function, even for the restricted case where $\beta = \mathbf{0}$. This appears to be a non-trivial problem, which in the renewal (i.i.d.) model (cf., Peña et al. (2001)) involves the baseline distribution renewal function owing to the sum-quota accrual scheme.

In the presence of model mis-specification, by examining the bias and rmse plots in Table 5, we find that under-specification leads to a non-negligible systematic bias that increases with n, and, based on other simulation runs not reported here, also with α. In fact, for this type of mis-specification, we have observed that the mean of the process $W_n(s^*, t)$ in t does not converge to the zero function as n increases, implying that with this mis-specification, the estimator $\hat{\bar{F}}_0$ may be inconsistent.

In contrast, with over-specification, we find that there is no recognizable loss in efficiency compared to the correct analysis, though we observe some very slight increase in the standard errors of the finite-dimensional parameter estimators (see XYZ-runs in Table 2 and compare the standard deviations in the A9 row of Table 1 and the X3 row, B9 row of Table 1 with the Y3 row, and the C9 row of Table 1 and the Z3 row). This indicates that there is much to be gained in the context of robustness by simply fitting the frailty-based model since, if the data did come from the frailty model, then the analysis is correct, while if the data came from the frailty-less model, there is no significant efficiency loss incurred; whereas, if there is under-specification of the model, then the consequences are unacceptable if the data actually came from the model with frailty. This lends strong support that this new class of models provides a general and flexible class for fitting recurrent event data and provides an avenue for a robust method of analysis for real data sets.

6 Applications to Real Data

In this section we will apply the estimation procedures developed in preceding sections to three real data sets from the biomedical and reliability settings.

The first application is to the bladder cancer data used in Wei, Lin, and Weissfeld (1989), which can be obtained from the `survival` package (Lumley and Therneau 2003) in the `R` Library. These data provide the times to recurrence of bladder cancer for $n = 85$ subjects. The covariates are X_1, the treatment indicator (1 = placebo; 2 = thiotepa); X_2, the size (in cm) of the largest initial tumor; and X_3, the number of initial tumors. A pictorial representation of the data is provided in the first plot in Figure 1, though the second and third covariates are not indicated. We first fitted the general model using the backward recurrence time $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$ as effective age. With $s^* = 64$, the maximum observation period, we fitted the general model *without* frailties, and obtained $\hat{\alpha} = 0.9826$ ($se = 0.0736$); $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (-0.3188, -0.0154, 0.1353)$. These are also the estimates obtained when the general model with frailty is fitted since in that case $\hat{\xi} = 5432999$ ($\hat{\eta} \approx 1$), very large value indicating that there is no need for the frailty component when the effective age is the backward recurrence time. Thus, using the approximate inverse of the partial likelihood information matrix from fitting the model without frailties, the associated estimated standard errors are .0736 for $\hat{\alpha}$ and $(0.2051, 0.0695, 0.0511)$ for $\hat{\beta}$. We recognize, however, that it remains to establish formally that these values are indeed valid standard error estimates, an issue to be addressed in a subsequent paper addressing the asymptotic properties of the model estimators.

For lack of information about the effective age, we also fitted the general model with frailties assuming a 'minimal repair' after each event, $\mathcal{E}(s) = s$. In this situation, the estimates are $\hat{\alpha} = .789$, $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (-.5743, -.0315, .2220)$, and $\hat{\xi} = .974$, indicating the importance of the frailty component in this case. The estimates of the survivor functions for the two effective age specifications are presented in the second plot in Figure 1. The lower curves (red), corresponding to the placebo group, are obtained by setting $X_1 = 1$ in the expression given by

$$\{\hat{\bar{F}}_0(t)\}^{\exp\{\hat{\beta}_1 X_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3\}},$$

while the upper curves (blue) are for the thiotepa group obtained by setting $X_1 = 2$. The observed means were $\bar{X}_2 = 2.01$ and $\bar{X}_3 = 2.11$. The solid curves are for the backward recurrence time effective age, while the dashed curves are for $\mathcal{E}(s) = s$. These plots seem to indicate that the thiotepa group has a higher survival rate than the placebo group, although the statistical significance of this difference depends on which effective age process was used.

It is of interest to compare the estimates of the regression coefficients from the general model with those obtained using the three existing methods of analysis described in Therneau and Hamilton (1997) and Therneau and Grambsch (2000). The table below summarizes the estimates from Andersen-Gill's (AG) method, Wei, Lin and Weissfeld's (WLW) marginal method, and Prentice, Williams and Peterson's (PWP) conditional method as reported in Therneau and Grambsch (2000), together with the estimates obtained from the general model with frailty under these two specifications of the effective age process, $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$ and $\mathcal{E}(s) = s$.

Covariate	Parameter	AG	WLW Marginal	PWP Conditional	General Model Perfect[a]	General Model Minimal[b]
$\log N(s-)$	α	-	-	-	.98 (.07)	.79
Frailty	ξ	-	-	-	∞	.97
rx	β_1	$-.47$ (.20)	$-.58$ (.20)	$-.33$ (.21)	$-.32$ (.21)	$-.57$
Size	β_2	$-.04$ (.07)	$-.05$ (.07)	$-.01$ (.07)	$-.02$ (.07)	$-.03$
Number	β_3	.18 (.05)	.21 (.05)	.12 (.05)	.14 (.05)	.22

[a]Effective Age is backward recurrence time ($\mathcal{E}(s) = s - S_{N^\dagger(s-)}$).
[b]Effective Age is calendar time ($\mathcal{E}(s) = s$).

From this table we note the crucial role that the effective age process plays in this analysis and how it provides a reconciliation of the varied estimates from these different methods. When the effective age process corresponds to perfect repair, then the estimates from the general model are close to those obtained from PWP's conditional method, whereas when the effective age process corresponds to minimal repair, the resulting estimates are close to those obtained from the WLW marginal method. The values from the AG method lie between these two cases. An explanation is that, in the analyses reported in Therneau and Grambsch (2000), the AG method and the WLW method assume a calendar time scale acting on the hazard functions, with the WLW method incorporating a stratification arising from the event occurrence, which seems to be modeled by the

α parameter in the general model; whereas, by the nature of the conditional approach of PWP, the time scale acting on the hazard is the backward recurrence time. The ability of the general model to seemingly explain these varied estimates from these different methods indicates the crucial role of the effective age and the need to monitor this information. Without this information, different methods will produce varied estimates, which could possibly lead to contradictory conclusions. If information about the effective age is obtained, then the general model may provide a flexible modeling vehicle for real data sets.

Because of the importance of the effective age process as demonstrated by this application to the bladder cancer data, we examined further through a simulation study the impact of mis-specifying the effective age process. We considered the model described in the simulation studies of Section 5 with a perfect repair probability of 0.6, and examined the impact of two types of effective age process mis-specification: that the interventions following event occurrences are all minimal repair, or that they are all perfect repair.

The results (not shown) indicate an interesting interplay between the nature of the baseline survivor function (DFR/IFR) and the behavior of $\hat{\bar{F}}_0$ and $\hat{\alpha}$. We observed that under the minimal repair mis-specification, when $\bar{F}_0$ is DFR, $\hat{\bar{F}}_0$ exhibits negative bias and $\hat{\alpha}$ is positively biased. Additionally for this mis-specification, when $\bar{F}_0$ is IFR, $\hat{\bar{F}}_0$ exhibits positive bias and $\hat{\alpha}$ is positively biased. Alternately, when the mis-specification is perfect repair, an underlying baseline DFR (IFR) is associated with positive (negative) bias in $\hat{\bar{F}}_0$ and negative (positive) bias in $\hat{\alpha}$. We explain these findings as follows: When the model mistakenly assumes minimal repair at each event occurrence, it tends to overestimate the effective age of units. Hence, in the case of DFR, the model anticipates longer interevent times than are realized in the data, creating the negative bias, especially for larger interevent times, in the estimates of the baseline survivor function in this situation. In the case of IFR, the minimal repair mis-specification leads to longer interevent times in the data than are anticipated by the model, creating a positive bias in the estimated baseline survivor function. When a perfect repair is incorrectly assumed at each event occurrence, the model tends to underestimate the effective age of units. Hence, using reasoning analogous to that for the minimal repair mis-specification, there is positive (negative) bias in the estimated baseline survivor function in the case of DFR (IFR). Especially interesting is that this behavior induces biases also in the finite-dimensional parameter estimates, with $\hat{\alpha}$, in particular, evidently compensating such that $\hat{\alpha}$ is positively biased when the baseline distribution is DFR, and negatively biased when this distribution is IFR. These simulation results further indicate the importance of monitoring the effective age process, and in future research we intend to examine consequences of other types of mis-specifications that may occur in practice.

Another biomedical example pertains to the rehospitalization of patients diagnosed with colorectal cancer. The data provide the calendar time (in days) of the successive hospitalizations after the date of surgery. The first readmission time was considered as the time between the date of the surgical procedure and the first rehospitalization after discharge related to colorectal cancer. Each subsequent readmission time was defined as the difference between the current hospitalization date and the previous discharge date. There were a total of 861 rehospitalization events recorded for the 403 patients included in the analysis. The data can be obtained from the *gcmrec* package in the `R` Library. The aim of the investigators was to determine whether there were differences regarding the time of the recurrent hospitalization due to social-demographic or clinical outcomes. However, in this example we consider only the following variables: tumor stage (Dukes classification: A-B, C or D); whether the patient received chemotherapy; and the distance between the hospital and the patient's residence. We have coded these covariates using dummy variables such that the regression coefficients can be interpreted as follows: β_1 pertains to patients diagnosed with Dukes C stage, and β_2 for patients with Dukes D stage; β_3 for patients who did not receive chemotherapy, and β_4 for patients whose residence are more than 30 kilometers from the hospital. Since in this case we have no information about the effective age, we assumed the backward recurrence time, $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$. We fitted the general model *without* frailties, taking $s^* = 2060$, the maximum follow-up time. The resulting estimates of the parameters are $\hat{\alpha} = 1.1243$ ($s.e. = 0.0145$), $\hat{\beta}_1 = 0.3102$ ($s.e. = 0.1204$), $\hat{\beta}_2 = 0.9270$ ($s.e. = 0.1369$), $\hat{\beta}_3 = -0.1226$ ($s.e. = 0.1062$), and $\hat{\beta}_4 = -0.0052$ ($s.e. = 0.148$). We also fitted the general model with frailties. After 35 iterations in the EM algorithm, the estimate of the frailty parameter ξ was quite small ($\hat{\xi} = 2.3934$), so we may conclude that the frailty component of the model is important for these data. The fitted frailty-based model provided the estimates: $\hat{\alpha} = 1.0811$, $\hat{\beta}_1 = 0.3050$, $\hat{\beta}_2 = 1.0516$, $\hat{\beta}_3 = -0.1426$, and $\hat{\beta}_4 = 0.0257$. Based on these results, we conclude that among these covariates, only the advanced tumor stages (C or D) are associated with an elevated risk of rehospitalization. Furthermore, since the estimate of α is larger than unity, there is an indication that each hospitalization increases the risk of further hospitalization.

The next data set, given in Blischke and Murthy (2000), which was analyzed in Kumar and Klefsjo (1992), concerns hydraulic load-haul-dump (LHD) subsystems used in moving ore and rock in underground mines in Sweden. The data set provides the calendar times (in hours), excluding repair or down times, of the successive failures of $n = 6$ such systems during the development phase, which was over a period of two years. We note that because in the data set the τ_i-values were not provided, for each unit we set $\tau_i = S_{iK_i}$. The first two machines are the oldest, the second two machines are of medium age, and the last two are relatively new machines. The categorized age of the machines will serve as our covariate, and it will be coded in terms of dummy variables with $\mathbf{X} = (0, 0)$ denoting old age, $\mathbf{X} = (1, 0)$ denoting medium age, and $\mathbf{X} = (0, 1)$ denoting young age. For purposes of our analysis, we will assume that the effective age function is the backward recurrence time $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$. The number of failure events for the six machines are $\mathbf{K} = (24, 26, 28, 29, 27, 24)$. When the general model without frailty is fitted, the resulting parameter estimates are $\hat{\alpha} = 1.0265$ and $(\beta_1, \beta_2) = (-0.0764, -0.0537)$. The estimates of the standard errors of these parameter estimates, obtained from the estimate of the inverse of the partial likelihood information matrix, are $\hat{\sigma}_{\hat{\alpha}} = 0.0106$ and $\hat{\sigma}_{\hat{\beta}} = (0.2014, 0.2056)$. These estimates were obtained by setting s^* to *any* value larger than $\max_{1 \le i \le 6} \tau_i = 4743$ hours. We also fitted the general model with gamma frailties to this hydraulic data set. The estimate of the frailty parameter was very large ($\hat{\xi} = 2.53 \times 10^{28}$ or $\hat{\eta} \approx 1$), which indicates the absence of unobserved frailties which would have induced additional heterogeneity among the machines. As a consequence, the estimates of α and (β_1, β_2) were identical to those obtained when the model without frailties was fitted.

7 Concluding Remarks

In this paper procedures for estimating the parameters of a general and flexible class of models for recurrent events proposed by Peña and Hollander (2004) were developed, and their properties were examined through computer simulation studies. The class of models, which includes as special cases many well-known models in survival analysis and reliability, possesses the appealing properties that it takes into account the effect of interventions which are administered after each event occurrence through the notion of an effective age, the possible weakening (or strengthening) effect of accumulating event occurrences, the possible presence of unobserved frailties that could be inducing correlations among the inter-event times per unit, and the effect of observable covariates. Some data sets in the biomedical and reliability/engineering settings were re-analyzed using this new class of models. It was found in the simulation studies that an under-specification of the model, in the sense of analyzing a data generated from the model with frailties using procedures developed from the model without frailties, could have unacceptable consequences in that the resulting estimators will have non-negligible systematic biases. On the other hand, it was found that over-specification of the model may provide a robust method of analysis with an acceptable loss in efficiency. The application of the procedures to the bladder cancer data set also provided a reconciliation of seemingly varied estimates obtained from currently available methods of analyzing recurrent event data, and highlights the importance of monitoring the effective age process.

There are still many interesting and important questions that need to be examined with regards to this general model. The first is the ascertainment of asymptotic properties of the estimators, such as their asymptotic normality or the weak convergence to a Gaussian process of a properly normed estimator of the baseline survivor function. This will be the topic of another paper, and the resolution of this asymptotic problem may require methods utilized in Murphy (1994, 1995) and Parner (1998). Some asymptotic results for specific models subsumed by the general class of models could be found in Peña et al. (2001) and Kvam and Peña (2003). Through such asymptotic analysis we will be able to obtain expressions for approximating analytically the standard errors of the estimators which will reflect the effects of an informative right-censoring mechanism as well as the impact of the sum-quota accrual scheme (see Peña et al. (2001) for the special case of a renewal model).

The problem of how to validate this class of models after it has been fitted to a specific data set is another open problem, and calls for suitable goodness-of-fit and model validation procedures. For example, in the illustration using the LHD data set, the survivor curve estimate for the medium age group is a little higher than for the new age group, and when one examines the data, there is a long gap in the third machine which might have led to this ordering. A question of interest is whether this particular inter-event time is an outlier, and it is hoped that future model validation and diagnostics procedures for this class of models will be able to answer such a question. Another question of interest is in the absence of an effective age data, might it have been better to fit a minimal repair effective age function, instead of the perfect repair effective age for this LHD data? This question leads to the recognition that an existing limitation of this class of models is that currently available data sets do not possess information regarding the effective age process. Thus, in applying this model to currently available data sets, we are forced to assume simple forms of the effective age process, such as the imperfect repair or perfect repair models discussed here. This problem of not knowing the effective age was first highlighted in Whitaker and Samaniego (1989), where they pointed out that if the repair modes, hence the effective ages, are not known in the minimal repair model, then the model is nonidentifiable. For the purpose of demonstrating their inference methods using Proschan (1963)'s air-conditioning data, which did *not* include the mode-of-repairs, they therefore augmented the inter-failure times data with assumed mode-of-repair data to illustrate the estimation of the reliability function. As demonstrated by our simulation studies to assess the impact of mis-specifying the effective age process in relation to the bladder cancer data application, a mis-specification on this effective age could lead to systematic biases on the estimators. It is therefore our hope that researchers will include the effective age in the data gathering stage of their studies. This may prove to be a novel and important aspect in areas in which recurrent events occur, and calls for a paradigm shift in the data gathering of recurrent event data.

References

Aalen, O. (1978), “Nonparametric inference for a family of counting processes,” *Annals of Statistics*, 6, 701–726.

Aalen, O. and Husebye, E. (1991), “Statistical analysis of repeated events forming renewal processes,” *Statistics in Medicine*, 10, 1227–1240.

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

Andersen, P. and Gill, R. (1982), “Cox's regression model for counting processes: a large sample study,” *Annals of Statistics*, 10, 1100–1120.

Baxter, L., Kijima, M., and Tortorella, M. (1996), “A point process model for the reliability of a maintained system subject to general repair,” *Stochastic Models*, 12, 37–65.

Blischke, W. and Murthy, P. (2000), *Reliability: Modeling, Prediction, and Optimization*, New York: Wiley-Interscience.

Block, H., Borges, W., and Savits, T. (1985), “Age-dependent minimal repair,” *J. Appl. Prob.*, 22, 51–57.

Brown, M. and Proschan, F. (1983), “Imperfect repair,” *J. Appl. Prob.*, 20, 851–859.

Cook, R. and Lawless, J. (2002), “Analysis of Repeated Events,” *Statistical Methods in Medical Research*, 11, 141–166.

Cox, D. (1972), “Regression models and life tables (with discussion),” *Journal of the Royal Statistical Society*, 34, 187–220.

— (1975), “Partial likelihood,” *Biometrika*, 62, 269–276.

Dempster, A., Laird, N., and Rubin, D. (1977), “Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion),” *J. Roy. Statist. Soc. B*, 39, 1–38.

Dorado, C., Hollander, M., and Sethuraman, J. (1997), “Nonparametric estimation for a general repair model,” *Ann. Statist.*, 25, 1140–1160.

Fleming, T. and Harrington, D. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.

Gail, M., Santner, T., and Brown, C. (1980), “An analysis of comparative carcinogenesis experiments based on multiple times to tumor,” *Biometrics*, 36, 255–266.

Gill, R. (1980), *Censoring and Stochastic Integrals*, Amsterdam: Mathematisch Centrum.

Gill, R. D. (1981), “Testing with replacement and the product-limit estimator,” *The Annals of Statistics*, 9, 853–860.

González, J., Slate, E., and Peña, E. (2003), *The gcmrec Package*, The Comprehensive R Archive Network, http://cran.r-project.org.

Hollander, M., Presnell, B., and Sethuraman, J. (1992), “Nonparametric methods for imperfect repair models,” *Ann. Statist.*, 20, 879–896.

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

Ihaka, R. and Gentleman, R. (1996), “R: a language for data analysis and graphics,” *Journal of Computational and Graphical Statistics*, 5, 299–314.

Jacod, J. (1975), "Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales," *Z. Wahrsch. verw. Geb*, 31, 235–253.

Jelinski, J. and Moranda, P. (1972), *Software reliability research*, New York: Academic Press, pp. 465–484.

Kijima, M. (1989), "Some results for repairable systems with general repair," *J. Appl. Prob.*, 26, 89–102.

Kumar, D. and Klefsjo, B. (1992), "Reliability analysis of hydraulic systems of LHD machines using the power law process model," *Reliability Engineering and System Safety*, 35, 217–224.

Kvam, P. and Peña, E. (2003), "Estimating Load-Sharing Properties in a Dynamic Reliability System," *Currently under revision for JASA*.

Last, G. and Szekli, R. (1998), "Asymptotic and monotonicity properties of some repairable systems," *Adv. in Appl. Probab.*, 30, 1089–1110.

Lawless, J. (1987), "Regression methods for Poisson process data," *J. Amer. Statist. Assoc.*, 82, 808–815.

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

Lumley, T. and Therneau, T. (2003), *The survival Package*, The Comprehensive R Archive Network, http://cran.r-project.org.

Murphy, S. (1994), "Consistency in a proportional hazards model incorporating a random effect," *The Annals of Statistics*, 22, 712–731.

— (1995), "Asymptotic theory for the frailty model," *The Annals of Statistics*, 23, 182–198.

Nielsen, G., Gill, R., Andersen, P., and Sorensen, T. (1992), "A counting process approach to maximum likelihood estimation in frailty models," *Scand. J. Statist.*, 19, 25–43.

Parner, E. (1998), "Asymptotic theory for the correlated gamma frailty model," *Annals of Statistics*, 26, 183–214.

Peña, E., Strawderman, R., and Hollander, M. (2000), *A Weak Convergence Result Relevant in Recurrent and Renewal Models*, Boston: Birkhauser, chap. 31, pp. 493–514.

Peña, E. and Hollander, M. (2004), "Models for Recurrent Events in Reliability and Survival Analysis," In *Mathematical Reliability* eds. T. Mazzuchi and N. Singpurwalla and R. Soyer, 105–123.

Peña, E., Slate, E., and Gonzalez, J. (2003), "Semiparametric Inference for a General Class of Models for Recurrent Events," Tech. Rep. 214, Department of Statistics, University of South Carolina.

Peña, E., Strawderman, R., and Hollander, M. (2001), "Nonparametric estimation with recurrent event data," *J. Amer. Statist. Assoc.*, 96, 1299–1315.

Prentice, R., Williams, B., and Peterson, A. (1981), "On the regression analysis of multivariate failure time data," *Biometrika*, 68, 373–379.

Presnell, B., Hollander, M., and Sethuraman, J. (1994), "Testing the minimal repair assumption in an imperfect repair model," *J. Amer. Statist. Assoc.*, 89, 289–297.

Proschan, F. (1963), "Theoretical explanation of observing decreasing failure rate," *Technometrics*, 5, 375–383.

Sellke, T. (1988), "Weak convergence of the Aalen estimator for a censored renewal process," In *Statistical Decision Theory and Related Topics IV* (eds., S. Gupta and J. Berger), 2, 183–194.

Stadje, W. and Zuckerman, D. (1991), "Optimal maintenance strategies for repairable systems with general degree of repair," *J. Appl. Prob.*, 28, 384–396.

Stocker, R. and Peña, E. (2003), "A General Class of Parametric Models for Recurrent Event Data," Currently under preparation.

Therneau, T. and Grambsch, P. (2000), *Modeling Survival Data: Extending the Cox Model*, New York: Springer.

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

Wang, M. C. and Chang, S. H. (1999), "Nonparametric estimation of a recurrent survival function," *Journal of the American Statistical Association*, 94, 146–153.

Wei, L., Lin, D., and Weissfeld, L. (1989), "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions," *J. Amer. Statist. Assoc.*, 84, 1065–1073.

Whitaker, L. and Samaniego, F. (1989), "Estimating the reliability of systems subject to imperfect repair," *J. Amer. Statist. Assoc.*, 84, 301–309.

	α	γ	ξ	η	n	NC	$\hat{\mu}_{Ev}$	$\hat{\alpha}$	$\hat{\sigma}_{\hat{\alpha}}$	$\hat{\beta}_1$	$\hat{\sigma}_{\hat{\beta}_1}$	$\hat{\beta}_2$	$\hat{\sigma}_{\hat{\beta}_2}$	$\hat{\eta}$
A2	0.9	0.9	2	0.67	30	0	4.1	0.898	0.031	1.012	0.379	−1.008	0.240	0.734
A3	0.9	0.9	2	0.67	50	0	5.2	0.899	0.021	1.017	0.287	−1.004	0.165	0.705
A5	0.9	0.9	6	0.86	30	0	4.3	0.900	0.030	0.988	0.300	−1.015	0.175	0.904
A6	0.9	0.9	6	0.86	50	0	5.3	0.899	0.021	0.998	0.221	−1.000	0.136	0.884
A8	0.9	0.9	∞	1.00	30	0	4.8	0.893	0.025	1.031	0.222	−1.030	0.135	
A9	0.9	0.9	∞	1.00	50	0	4.4	0.895	0.018	1.024	0.158	−1.023	0.104	
A11	0.9	2.0	2	0.67	30	0	7.8	0.902	0.016	1.010	0.348	−1.018	0.202	0.721
A12	0.9	2.0	2	0.67	50	0	6.7	0.902	0.012	0.994	0.271	−1.012	0.144	0.710
A14	0.9	2.0	6	0.86	30	0	8.9	0.900	0.016	1.009	0.236	−1.008	0.135	0.895
A15	0.9	2.0	6	0.86	50	0	7.2	0.900	0.012	0.998	0.173	−1.004	0.101	0.882
A17	0.9	2.0	∞	1.00	30	0	8.4	0.898	0.015	1.017	0.155	−1.014	0.095	
A18	0.9	2.0	∞	1.00	50	0	7.4	0.899	0.011	1.003	0.112	−1.007	0.072	
B2	1	0.9	2	0.67	30	2	9.5	1.000	0.011	1.010	0.374	−1.000	0.227	0.735
B3	1	0.9	2	0.67	50	0	8.7	1.000	0.007	0.989	0.280	−1.002	0.165	0.704
B5	1	0.9	6	0.86	30	0	7.7	1.000	0.012	1.014	0.286	−0.993	0.164	0.901
B6	1	0.9	6	0.86	50	0	7.3	1.000	0.007	1.013	0.201	−0.999	0.118	0.880
B8	1	0.9	∞	1.00	30	0	8.1	0.998	0.008	1.029	0.185	−1.024	0.114	
B9	1	0.9	∞	1.00	50	0	9.1	0.999	0.006	1.010	0.130	−1.012	0.084	
B11	1	2.0	2	0.67	30	0	9.5	1.000	0.008	1.016	0.336	−1.028	0.194	0.725
B12	1	2.0	2	0.67	50	0	13.0	1.000	0.006	1.004	0.258	−1.012	0.146	0.705
B14	1	2.0	6	0.86	30	0	13.8	1.000	0.008	1.006	0.228	−1.002	0.132	0.889
B15	1	2.0	6	0.86	50	0	10.8	1.000	0.006	1.003	0.168	−1.001	0.097	0.876
B17	1	2.0	∞	1.00	30	0	14.0	0.999	0.007	1.017	0.133	−1.010	0.083	
B18	1	2.0	∞	1.00	50	0	11.2	1.000	0.005	1.010	0.099	−1.006	0.065	
C2	1.05	0.9	2	0.67	30	3	11.8	1.051	0.007	0.994	0.366	−0.994	0.222	0.730
C3	1.05	0.9	2	0.67	50	0	9.7	1.050	0.004	1.009	0.284	−0.993	0.153	0.703
C5	1.05	0.9	6	0.86	30	1	12.9	1.051	0.007	1.002	0.271	−0.993	0.160	0.899
C6	1.05	0.9	6	0.86	50	0	13.9	1.050	0.005	1.006	0.196	−0.992	0.119	0.880
C8	1.05	0.9	∞	1.00	30	0	10.9	1.049	0.007	1.020	0.154	−1.012	0.101	
C9	1.05	0.9	∞	1.00	50	0	13.8	1.050	0.004	1.009	0.121	−1.006	0.072	
C11	1.05	2.0	2	0.67	30	0	12.3	1.050	0.006	1.026	0.336	−1.018	0.184	0.726
C12	1.05	2.0	2	0.67	50	0	13.4	1.050	0.005	1.008	0.248	−1.012	0.136	0.705
C14	1.05	2.0	6	0.86	30	0	10.9	1.050	0.006	1.019	0.225	−1.000	0.124	0.890
C15	1.05	2.0	6	0.86	50	0	14.3	1.050	0.004	0.997	0.166	−1.000	0.096	0.876
C17	1.05	2.0	∞	1.00	30	0	18.5	1.050	0.005	1.004	0.123	−1.010	0.076	
C18	1.05	2.0	∞	1.00	50	0	13.5	1.050	0.004	1.004	0.090	−1.003	0.054	

Table 1: Summary of simulated means and standard deviations of the estimators of α, β, and $\eta = \xi/(1+\xi)$. The true value of β is $(1, -1)$, and 1000 replications were run for each parameter combination. The other columns of this table are: γ denotes the Weibull shape parameter; n is the sample size; NC is the number of replicates in which there was no convergence; $\hat{\mu}_{Ev}$ is the observed mean number of events per unit in all the simulation replications.

	α	γ	ξ	n	NC	$\hat{\mu}_{\hat{\alpha}}$	$\hat{\sigma}_{\hat{\alpha}}$	$\hat{\mu}_{\hat{\beta}_1}$	$\hat{\sigma}_{\hat{\beta}_1}$	$\hat{\mu}_{\hat{\beta}_2}$	$\hat{\sigma}_{\hat{\beta}_2}$
U2	0.90	0.9	2	30	0	0.954	0.031	0.779	0.322	-0.770	0.210
U3	0.90	0.9	2	50	0	0.959	0.023	0.747	0.239	-0.740	0.154
U5	0.90	0.9	6	30	0	0.921	0.028	0.898	0.285	-0.919	0.168
U6	0.90	0.9	6	50	0	0.923	0.020	0.883	0.212	-0.888	0.131
U8	0.90	2.0	2	30	0	0.952	0.022	0.719	0.297	-0.728	0.187
U9	0.90	2.0	2	50	0	0.956	0.017	0.700	0.223	-0.707	0.139
U11	0.90	2.0	6	30	0	0.920	0.018	0.909	0.220	-0.901	0.138
U12	0.90	2.0	6	50	0	0.922	0.013	0.879	0.167	-0.879	0.101
V2	1.00	0.9	2	30	0	1.019	0.014	0.771	0.324	-0.751	0.215
V3	1.00	0.9	2	50	0	1.020	0.009	0.726	0.251	-0.715	0.157
V5	1.00	0.9	6	30	0	1.008	0.011	0.913	0.271	-0.888	0.172
V6	1.00	0.9	6	50	0	1.009	0.008	0.886	0.198	-0.868	0.129
V8	1.00	2.0	2	30	0	1.024	0.012	0.711	0.291	-0.723	0.191
V9	1.00	2.0	2	50	0	1.024	0.009	0.685	0.221	-0.695	0.136
V11	1.00	2.0	6	30	0	1.009	0.009	0.885	0.224	-0.879	0.137
V12	1.00	2.0	6	50	0	1.009	0.008	0.871	0.173	-0.867	0.104
W2	1.05	0.9	2	30	0	1.059	0.010	0.725	0.342	-0.720	0.226
W3	1.05	0.9	2	50	0	1.058	0.006	0.696	0.261	-0.691	0.158
W5	1.05	0.9	6	30	0	1.054	0.007	0.873	0.273	-0.869	0.179
W6	1.05	0.9	6	50	0	1.053	0.005	0.851	0.198	-0.842	0.128
W8	1.05	2.0	2	30	0	1.061	0.009	0.704	0.296	-0.704	0.179
W9	1.05	2.0	2	50	0	1.062	0.007	0.686	0.224	-0.684	0.138
W11	1.05	2.0	6	30	0	1.054	0.007	0.877	0.227	-0.880	0.132
W12	1.05	2.0	6	50	0	1.054	0.005	0.870	0.162	-0.870	0.104
X2	0.90	0.9	∞	30	0	0.893	0.026	1.030	0.224	-1.031	0.144
X3	0.90	0.9	∞	50	2	0.895	0.018	1.030	0.173	-1.022	0.105
X5	0.90	2.0	∞	30	0	0.897	0.015	1.016	0.163	-1.015	0.099
X6	0.90	2.0	∞	50	2	0.898	0.011	1.014	0.115	-1.015	0.076
Y2	1.00	0.9	∞	30	6	0.998	0.010	1.023	0.186	-1.022	0.116
Y3	1.00	0.9	∞	50	1	0.999	0.006	1.019	0.136	-1.019	0.086
Y5	1.00	2.0	∞	30	2	0.999	0.007	1.013	0.138	-1.011	0.084
Y6	1.00	2.0	∞	50	0	0.999	0.006	1.010	0.100	-1.007	0.066
Z2	1.05	0.9	∞	30	6	1.050	0.005	1.015	0.162	-1.011	0.099
Z3	1.05	0.9	∞	50	6	1.050	0.004	1.017	0.112	-1.013	0.073
Z5	1.05	2.0	∞	30	2	1.050	0.005	1.014	0.126	-1.008	0.078
Z6	1.05	2.0	∞	50	2	1.050	0.004	1.005	0.094	-1.005	0.055

Table 2: Summary of simulated means and standard deviations for the estimators of α, β_1, and β_2 for the situation of under-specification (label UVW) and over-specification (label XYZ). The true regression coefficients are $\beta = (1, -1)$ and 1000 replications were run for each parameter combination.

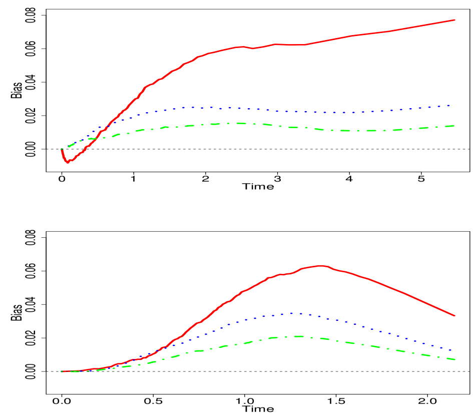
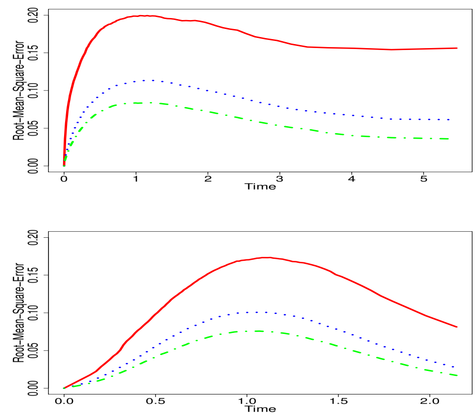
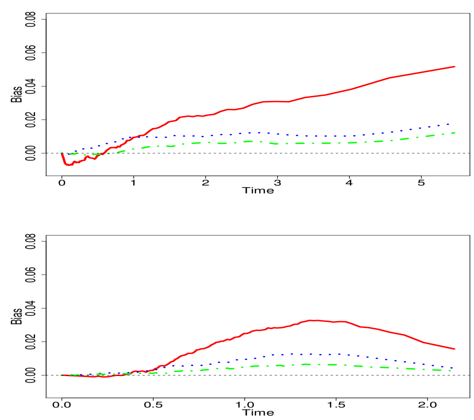
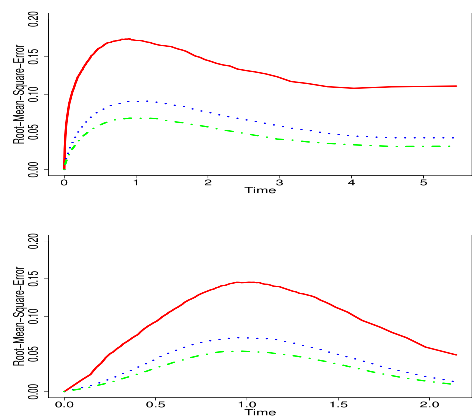
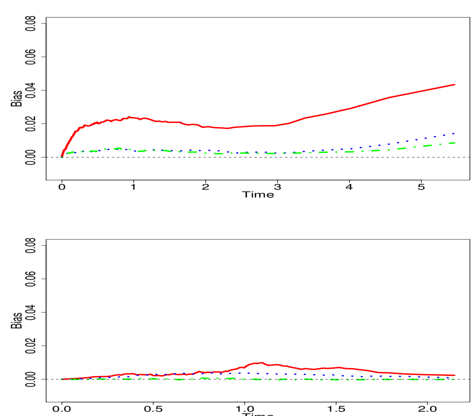
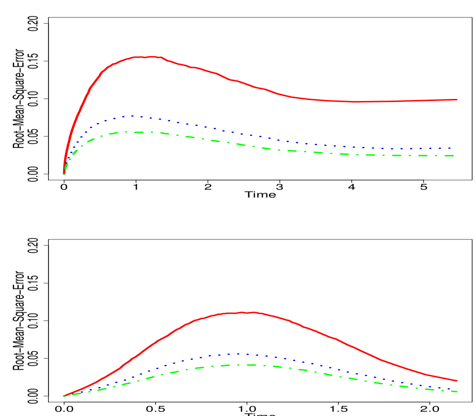
ξ	Bias	RMSE
2		
6		
∞		

Table 3: Bias and root mean squared error curves for the estimator of the baseline survivor function as the sample size (n) varies [$n = 10$ is red; $n = 30$ is blue; $n = 50$ is green]. This is for the case where $\alpha = .90$. The upper plot frame in each cell is for Weibull shape parameter of 0.90, while the lower plot frame is for shape parameter of 2.0.

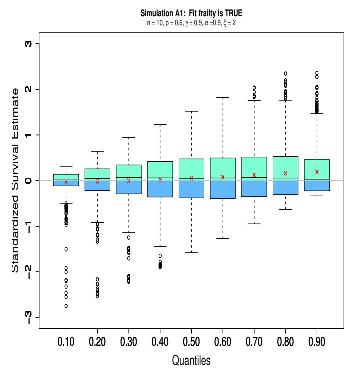
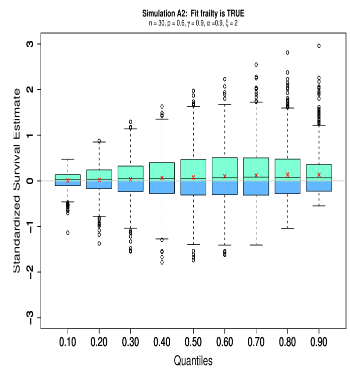
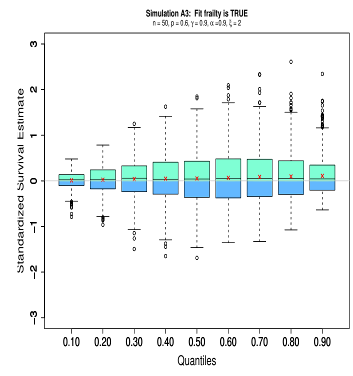
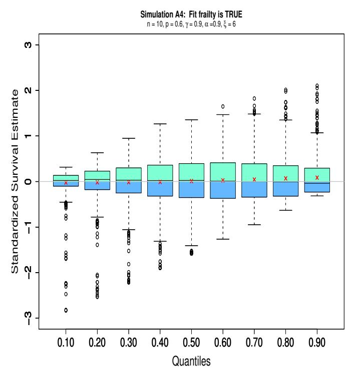
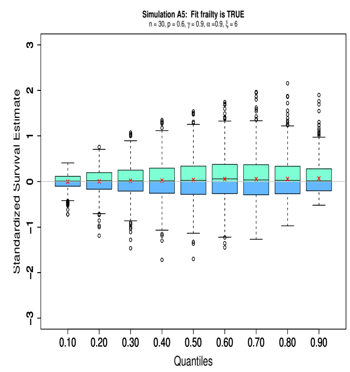
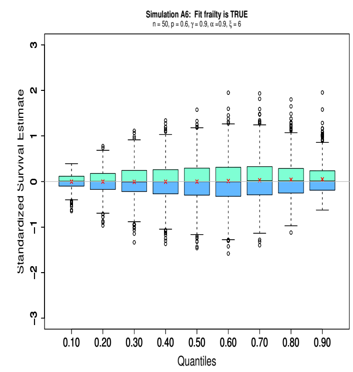
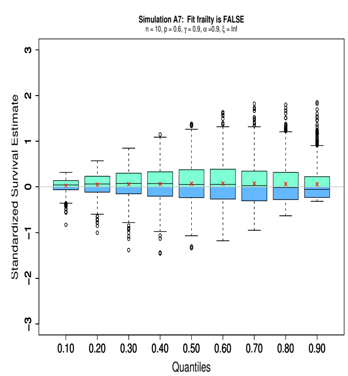
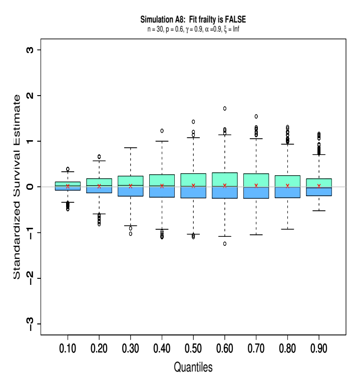
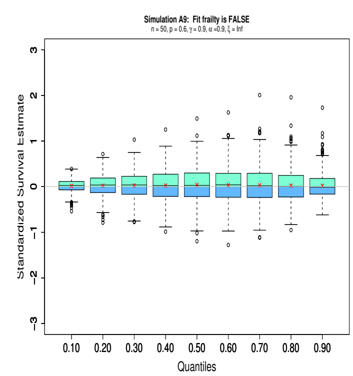
<table>
<tr><th>ξ</th><th>$n = 10$</th><th>$n = 30$</th><th>$n = 50$</th></tr>
<tr><td>2</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>6</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>∞</td><td>
</td><td>
</td><td>
</td></tr>
</table>

Table 4: Side-by-side boxplots of the simulated values of $W_n(s^*, t) = \sqrt{n}[\hat{\bar{F}}_0(s^*, t) - \bar{F}_0(t)]$ for t-values associated with the $[.1 : (.1) : .9]$ percentiles of the true baseline survivor function $\bar{F}_0$. The plots are for the case where $\bar{F}_0$ is a Weibull with shape parameter of 0.9 and scale parameter of unity, and with $\alpha = .90$, and for combinations of ξ and n as indicated. In each boxplot, the mean value is indicated by a red $\times$, black line marks the median, and darker shade (lighter shade) indicates below (above) zero.

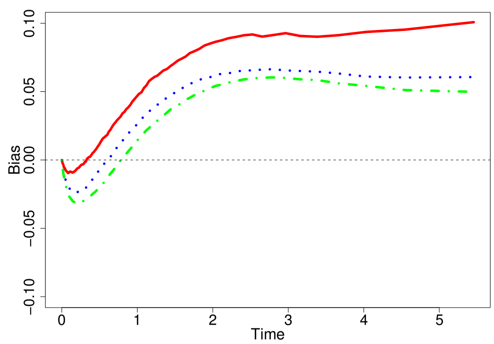
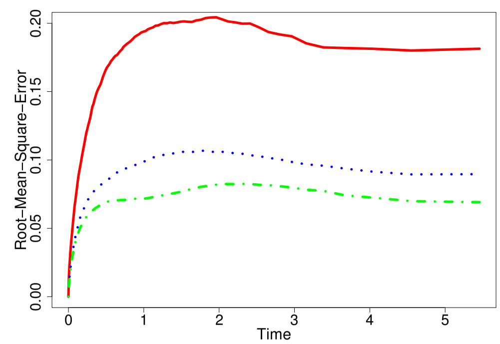
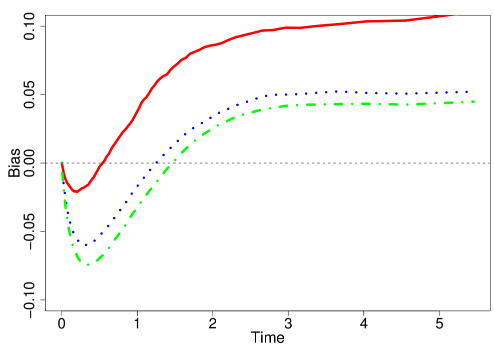
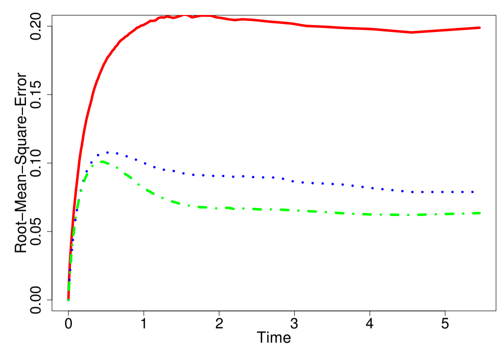
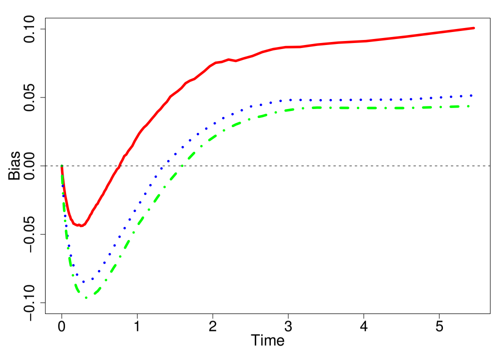
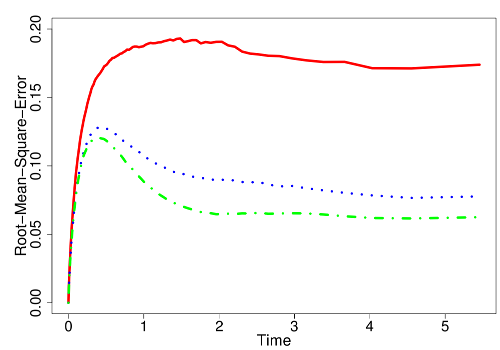
α	Bias	RMSE
.9		
1		
1.05		

Table 5: Bias and RMSE curves when data are generated from the model with frailty with $\xi = 2$, $\gamma = 0.9$, and the model with no frailty is fitted. The three curves are for varying sample sizes with red curve for $n = 10$, blue curve for $n = 30$, and green curve for $n = 50$.

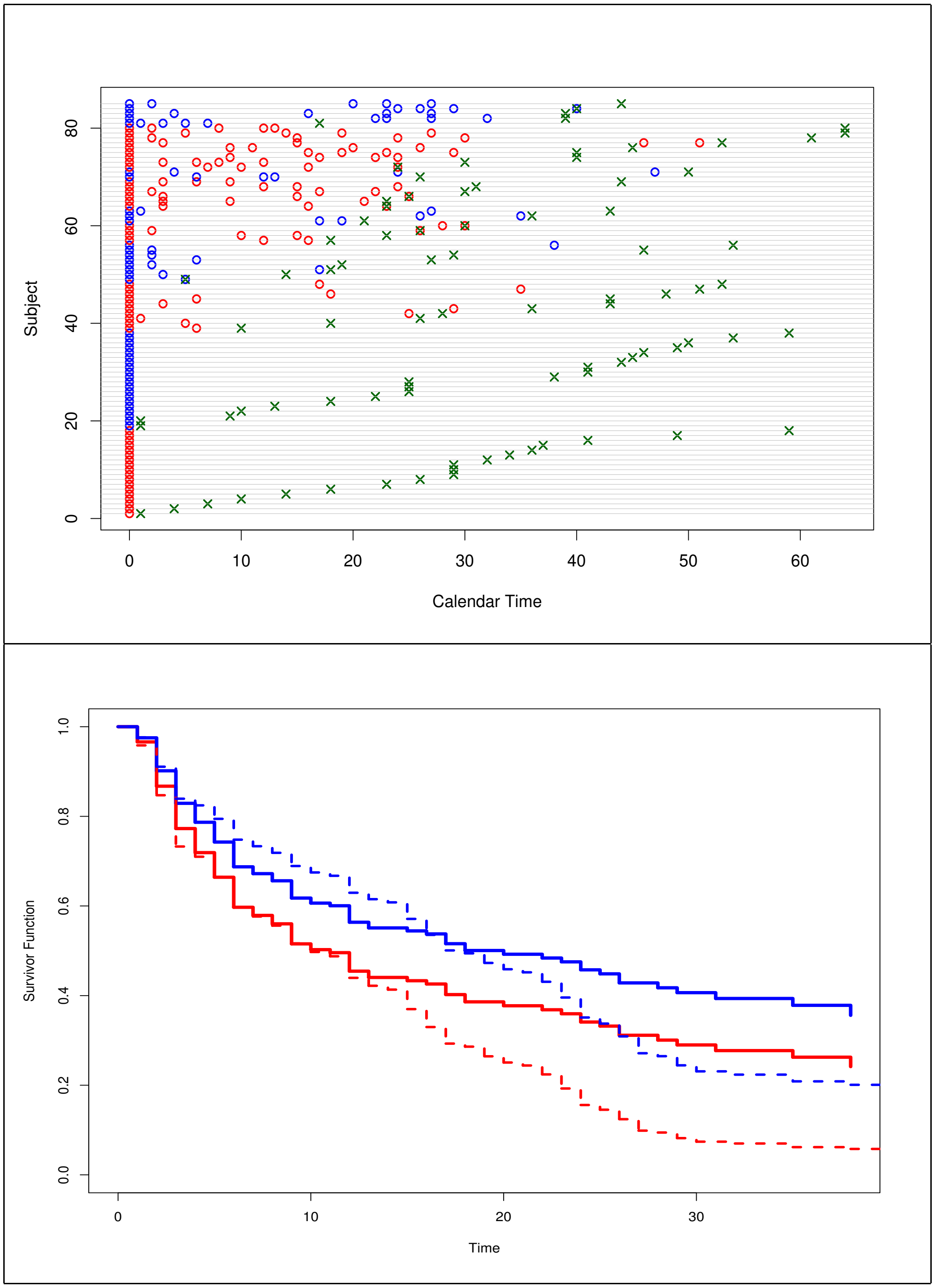


Figure 1: The first plot is a pictorial representation of the bladder data set used by Wei, Lin and Weissfeld (1989). The picture shows the times of bladder cancer recurrence for 85 subjects. The treatment assignment (X_1) is color-coded according to red (placebo) and blue (thiotepa). The other two covariates, the size of the largest initial tumor (X_2) and the number of initial tumors (X_3) are not depicted in this picture. The second plot contains estimates of the survivor function for this data set when the model without (and with) frailty is fitted. The red curve is for the placebo group ($X_1 = 1$), while the blue curve is for the thiotepa group ($X_1 = 2$), both evaluated at the mean values of X_2 and X_3. The solid curves are for effective age $\mathcal{E}(s) = s - S_{N^{\dagger}(s-)}$ (perfect repair), while the dashed curves are when $\mathcal{E}(s) = s$ (minimal repair).