Statistical Methodology 10 (2013) 1–13

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Semiparametric inference with correlated recurrence time data

Akim Adekpedjou [a,*], Jonathan Quiton [b], Xuerong Meggie Wen [a]

[a] *Department of Mathematics and Statistics, Missouri University of Science and Technology (formerly University of Missouri-Rolla), Rolla, MO 65409, United States*

[b] *Department of Mathematics and Computer Science, Western Kentucky University, Bowling Green, KY 42101, United States*

ARTICLE INFO

Article history:
Received 15 January 2011
Received in revised form
6 September 2011
Accepted 20 May 2012

Keywords:
Correlated recurrence times
Informative monitoring
Frailty model
Martingales
EM algorithm

ABSTRACT

We consider a study which monitors the occurrences of a recurrent event for n subjects or units. Recurrent event data have many features which are worth looking into in the estimation process. In this manuscript, we consider the problem of estimating the distribution function of the inter-event times by taking into account two of these features: correlation among the inter-event times and the dependence and informative aspect of the right-censoring random variables. The parametric approach to the problem has been dealt with in Zamba and Adekpedjou (2011) [25]. The semiparametric approach is considered in this article. We derive a Kaplan–Meier type estimator of the distribution function under the gamma frailty model and an informative monitoring model for recurrent events by extending an approach due to Sellke (1988) [20]. The sampling distribution properties of the proposed estimators are examined through simulation studies. Furthermore, the performance of our proposed estimator is assessed with respect to the existing ones. The procedures are applied to a recurrent event dataset.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Recurrent event data is a multivariate lifetime data where the event of interest recurs a random number of times over a random observation window. Recurrent events are observed in many disciplines such as biomedical, economic, engineering, actuarial science, and social sciences applications

* Corresponding author. Tel.: +1 5733414649; fax: +1 5733414647.
E-mail addresses: akima@mst.edu (A. Adekpedjou), Jonathan.Quiton@wku.edu (J. Quiton), wenx@mst.edu (X.M. Wen).

1572-3127/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2012.05.001

to name but a few. Examples of such events include recurrent hospitalization of patients with chronic disease, repeated failure of machines in engineering, habitual claim filing in insurance, and recurrent economic recessions. It is therefore of interest to develop stochastic models for analyzing recurrence behavior. Recognizing this need, Peña and Hollander [16] introduced a flexible class of models for analyzing these types of data. Their proposed models incorporate most of the different aspects of recurrent event data, such as the impacts of event accumulation, the effects of covariates, the impacts of performed intervention after each event occurrence, the impacts of informative and dependent censoring, and the impacts of latent or unobserved variables which induce a correlation among inter-event times.

Estimation of the distribution function governing the time to occurrence of a recurrent event has been widely studied over the past several decades [19,24,6,7,23,22]. For instance, [19], assuming the inter-event times represent an independent and identically distributed (i.i.d.) sample from a fixed distribution, propose a generalization of the Kaplan–Meier estimator (cf. [10]) and showed that it has the nonparametric maximum likelihood estimator property under the i.i.d. assumption on the gap time. Furthermore, they show that their estimator is inconsistent in the case where the inter-occurrence times satisfy a multiplicative frailty model. Their proposed estimator does not account for the effect of informative monitoring. Wang and Chang [24] propose an estimator that is valid for both i.i.d. and correlated inter-event times, but their estimator ignores informative monitoring (a point to which we return later) and accounts for the right-censoring random variables only when a unit does not have a complete observation. Consequently, information is ignored. This work proposes and investigates the properties of an estimator that utilizes all the right-censored observations on all subjects whether or not a unit has a complete observation. Other works in the area include that of [12,8]. In [12], the dependence between events and risk of death was modeled using a shared frailty model. In [8], association between the recurrent events process and the failure time was modeled with a common subject-specific latent variable. For a detailed review of literature on the topic, we refer the reader to [4].

Two very important features of recurrent event data are: (1) informativeness and dependence of the right-censoring random variable and (2) a possible correlation among inter-event times. These two topics are discussed, respectively. Denote by $T_{ij} = S_{ij} - S_{i,j-1}, i = 1, \ldots, n; j = 1, 2, \ldots$ the inter-event times, where S_{ij} are the calendar times, with $S_{i0} = 0$. The T_{ij}s are assumed to be i.i.d. (for now) with distribution function, F. K_i is the random number of event recurrences observed over the observation window $[0, \tau_i]$. Therefore, the random observable for subject i is represented by the vector $\mathbf{D}_i = (K_i, \tau_i, T_{i1}, \ldots, T_{iK_i}, \tau_i - S_{iK_i})$. The vector $(T_{i1}, \ldots, T_{iK_i}, T_{iK_i+1})$ satisfies the sum-quota constraint given by $\sum_{j=1}^{K_i} T_{ij} \leq \tau_i < \sum_{j=1}^{K_i} T_{ij} + T_{i,K_i+1}$. Hence, the random observation window $[0, \tau_i]$ is informative of the inter-event times distribution. Furthermore, the time after which they have been censored defines the duration of their monitoring. Therefore, the recurrent event data accrual scheme has both an informative censoring mechanism as well as a dependent censoring structure which we refer to as *informative monitoring*. In the single event setting, *informative censoring* occurs when the observation window is terminated by something other than the event of interest, usually called competing risks. Note that the two concepts coincide in the single event setting. Now, the assumption that the T_{ij}s are i.i.d. is sometimes a restrictive assumption, especially in biomedical studies where the T_{ij}s are more likely to be correlated. This correlation induces a random effect component inherent to each subject. If these two features are not properly accounted for, erroneous conclusions can be made on the estimation of model parameter. [5,21,17,4].

To account for informative monitoring, we consider the estimation of the distribution function of the inter-event times F when there is an additional structure in the model through a relationship between F and G, where G is the distribution of τ. The relationship postulates that the existence $\beta > 0$ such that $1 - G = (1 - F)^{\beta}$. This model was introduced in [11] for the single event setting and is well known in the literature as the Koziol–Green (KG) model. In the random censorship model (single event setting), β is referred to as the censoring parameter since the probability of being censored under the model is $\beta/(1 + \beta)$. The KG model for recurrent events was introduced in [4] and was referred to as the generalized KG model. Furthermore, β was referred to as the monitoring parameter since it determines the length of the monitoring period relative to the inter-event times. We acknowledge that the KG model is a strong assumption connecting monitoring time and gap time distributions. However, just as in the single event setting, the utility of the generalized KG model is not primarily to provide a practical and realistic model, but rather to provide a medium in which to examine analytical properties of inference procedures with recurrent event data. To model the association among the inter-event times, we use the *frailty* model. Frailties are unobservable random effect components that are used in survival analysis. A specific model for correlated recurrence times is the multiplicative frailty model which is described in detail in Section 2.1.

The joint analysis of correlated recurrence time and informative monitoring when the distribution function of the inter-event time is assumed to be fully parametric has been considered in [25]. The current manuscript deals with the semiparametric approach and aims to develop a semiparametric estimator of the inter-event times distribution under the generalized KG and multiplicative frailty models. Furthermore, the performance of the proposed estimator with respect to a fully nonparametric estimator in [19] as well as the estimator in [3] is assessed. The main contribution of this manuscript is the introduction of informative monitoring in recurrent event models with frailty. This is in addition to the right-censored gap times which are always informative of $T_{i,K_{i+1}}$.

We now outline the contents of this manuscript. In Section 2, we describe the multiplicative gamma frailty model, the random entities and the relevant stochastic processes which are used to derive the proposed estimator in Section 2.3. We adopt the framework of analyzing failure time data introduced by Aalen [1] using counting processes and martingale theory. In Section 3, we outline steps for estimating the frailties and the model parameters. Finally, Sections 4 and 5 summarize the results of simulation studies, and illustrate the estimation procedures using the gastroenterology data of [2].

2. Semiparametric estimation

2.1. Correlated recurrence time data and random entities

Consider n experimental units with the i^{th} unit monitored for occurrence of a recurrent event over a period $[0, \tau_i]$. The τ_i's are right-censoring random variables, i.i.d. with distribution function $G(t) = P(\tau_i \le t)$ and survivor function $\bar{G} \equiv 1 - G$. For each unit i, observe events at calendar times $0 = S_{i0} < S_{i1} < \cdots$ with $K_i = \max\{k \in \mathbb{N} : S_{ik} \le \tau_i\}$ the total number of event occurrences per unit. Denote by $T_{ij} \equiv S_{ij} - S_{ij-1}$ the successive inter-event times. These are assumed to represent an i.i.d. sample from the distribution function F_0. Recall, the random observable for the i^{th} subject is

$$\mathbf{D}_i = (K_i, \tau_i, T_{i1}, \ldots, T_{iK_i}, \tau_i - S_{iK_i}), \quad i = 1, 2, \ldots, n. \tag{1}$$

To describe correlated recurrence times, we assume that there exists a per unit random variable Z_i, i.i.d. with distribution function $H(\cdot|\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Theta \subseteq \Re^p$. If the Z_is are known quantities, then the T_{ij}s are assumed to be i.i.d. with distribution function F having a hazard rate and survivor function given by

$$\lambda_F(t|Z_i) = Z_i \lambda_0(t) \quad \text{and} \quad \bar{F}(t|Z_i = z_i) = \{\bar{F}_0(t)\}^{z_i} = \{\exp(-\Lambda_0(t))\}^{z_i}, \tag{2}$$

respectively, where $\lambda_0(t)$ is the baseline hazard rate function, $\Lambda_0(t)$ is the cumulative baseline hazard rate function, and $\bar{F}_0(t)$ is the baseline survivor function. However, since the Z_is are random quantities, they will be estimated using Eq. (15). The frailty distribution considered in this work is a gamma distribution with equal scale and shape parameters α. The parameter α controls the degree of association among the inter-event times. In particular, if $\alpha \to \infty$, we obtain the estimator in the i.i.d. setting. The equal parameters restriction is imposed to ensure model identifiability. Thus, under the gamma frailty model, the common conditional survivor function is given by

$$\bar{F}(t|Z = z) = \{\bar{F}_0(t)\}^z = \{\exp(-\Lambda_0(t))\}^z = \left[\frac{\alpha}{\alpha + \Lambda_0(t)}\right]^{\alpha}. \tag{3}$$

Note that the observable random vector for subject i with correlated recurrence times is still the observable in (1) since the Z_is are not observed.

2.2. Stochastic process formulation

For more details on what follows, we refer the reader to [18,19]. $(\Omega, \mathcal{F}, P)$ is the probability space on which all random entities are defined. For $i = 1, 2, \ldots, n; j = 1, 2, \ldots,$ and $s \in \Re$, let the calendar time processes be defined by

$$N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \le s \wedge \tau_i\}, \qquad Y_i^\dagger(s) = I\{\tau_i \ge s\}, \quad \text{and} \quad N_i^\tau(s) = I\{\tau_i \le s\}. \tag{4}$$

The processes in (4) counts the number of failures for unit i; indicates whether unit i is still under observation at time s; and indicates if unit i is past its monitoring time at time s, respectively. We augment our probability space by the natural filtration $\mathbf{F} = \{\mathcal{F}_s : s \ge 0\}$, where $\mathcal{F}_s$ is the history of the process up to time s. For each s, define the backward recurrence time via $R_i(s) = s - S_{iN_i^\dagger(s-)}$. This expression indicates the length of time elapsed since the last event occurrence. This is an $\mathbf{F}$-adapted and left-continuous process, hence $\mathbf{F}$-predictable. Assuming the generalized KG model holds, notice that the hazard functions of τ_i and T_{ij} under the correlated recurrent event data are given by $\Lambda_G(\cdot) = \beta z \Lambda_0(\cdot)$ and $\Lambda_F(\cdot) = z\Lambda_0(\cdot)$, respectively. If the frailty model holds, conditional on $Z_i = z_i$, let

$$A_i^\dagger(s|z_i) = \int_0^s Y_i^\dagger(w) z_i \lambda_0(R_i(w))dw \quad \text{and} \quad A_i^\tau(s|z_i) = \int_0^s Y_i^\dagger(w)\beta z_i \lambda_0(w)dw. \tag{5}$$

From stochastic integration theory, $A_i^\dagger(s|z_i)$ and $A_i^\tau(s|z_i)$ are the compensators of $N_i^\dagger(s)$ and $N_i^\tau(s)$, respectively. Therefore,

$$M_i^\dagger(s|z_i) = N_i^\dagger(s) - A_i^\dagger(s|z_i) \quad \text{and} \quad M_i^\tau(s, \beta|z_i) = N_i^\tau(s) - A_i^\tau(s, \beta|z_i) \tag{6}$$

are, for each i, square-integrable martingales with respect to the filtration $\mathbf{F}$. Observe that the first process in (6) is a calendar time process. For estimating $\Lambda_0(t)$, gap time processes should be considered. Since the calendar time processes involve the random argument $R_i(w)$, the usual martingale theory can not be directly applied. As such, the processes in (6) will be transformed by extending the idea of [20] to obtain processes that make the connection between calendar time and gap time, thereby enabling the use of martingale methods. Those transformations are described below.

2.3. Method of moment estimation

Let $(s, t) \in [0, \infty)^2$, where s represents calendar time and t represents gap time. Define the process $Z_i(s, t) = I(R_i(s) \le t)$ which indicates whether at most t units of time has elapsed since the last event recurrence. Following [18], define the doubly-indexed processes as

$$N_i(s, t) = \int_0^s Z_i(w, t)dN_i^\dagger(w),$$

$$A_i(s, t|z_i) = \int_0^s Z_i(w, t)dA_i^\dagger(w|z_i),$$

$$M_i(s, t|z_i) = \int_0^s Z_i(w, t)dM_i^\dagger(w|z_i).$$

It can be shown that $A_i(s, t|z_i) = \int_0^t Y_i(s, w)z_i\lambda_0(w)dw$ (cf. Proposition 1 of [18]), where

$$Y_i(s, t) = \sum_{j=1}^{N_i^\dagger(s-)} I\{T_{ij} \ge t\} + I\left\{(s \wedge \tau_i) - S_{iN_i^\dagger(s-)} \ge t\right\}. \tag{7}$$

The process $N_i(s, t)$ represents the number of events that have occurred on or before calendar time s for the ith subject and whose inter-event times are at most t, whereas the process $Y_i(s, t)$ represents the generalized at-risk process. Furthermore, since the components of $Y_i(s,t)$ are i.i.d. processes, we have (cf. Proposition 2, [19])

$$\sup_{s\in[0,t^\star]}\left|\frac{1}{n}\sum_{i=1}^{n}Y_i(s,t)-y(s,t)\right|=o_p(1),$$

as $n\to\infty$, where $y(s,t)=E\{Y_1(s,t)\}=\bar{F}(t)\bar{G}(t)+\bar{F}(t)I\{t\le s\}\int_t^s\rho(w-t)dG(w)$ and $\rho(t)$ is the renewal function of F_0. Using Proposition 1 of [18], we have the identity $M_i(s,t|z_i)=N_i(s,t)-A_i(s,t|z_i)=N_i(s,t)-\int_0^t Y_i(s,w)z_i\lambda_0(w)dw$. Letting $\mathbf{z}=(z_1,\ldots,z_n)$, we define the aggregated processes

$$M(s,t|\mathbf{z})=\sum_{i=1}^{n}M_i(s,t|z_i)\quad\text{and}\quad M^\tau(t,\beta|\mathbf{z})=\sum_{i=1}^{n}M_i^\tau(t,\beta|z_i).$$

It follows that, for a given study period $[0,s]$ and a fixed gap time t, a method of moments estimator of $\Lambda_0(t)$, given the frailties $z_1,\ldots,z_n$, denoted by $\hat{\Lambda}_0(s,t|\hat{\beta},z_1,\ldots,z_n)$ is given by

$$\hat{\Lambda}_0(s,t|\hat{\beta};\mathbf{z})=\int_0^t J(s,w;\hat{\beta}|z_i)\left\{\frac{N(s,dw)+N^\tau(dw)}{\sum_{i=1}^{n}z_i(Y_i(s,w)+\hat{\beta}Y_i^\tau(w))}\right\},\tag{8}$$

where $J(s,w;\hat{\beta}|z_i)=I\{z_i(Y_i(s,w)+\hat{\beta}Y_i^\tau(w)>0)\}$ is a bounded left-continuous predictable process and $\hat{\beta}$ is the maximum likelihood estimate of β. The maximum likelihood estimate of β is obtained using a procedure we will describe in Section 4. Note that $\hat{\Lambda}_0(s,t|\hat{\beta};\mathbf{z})$ is a semiparametric maximum likelihood estimator and $\lim_{s\to\infty}\hat{\Lambda}_0(s,t|\hat{\beta};\mathbf{z})=\hat{\Lambda}_0(t|\hat{\beta};\mathbf{z})$. Moreover, the estimator is a generalized version of the estimator obtained in Section 4.2 of [19]. As $z_i\xrightarrow{p}1$ (here $\xrightarrow{p}$ denotes convergence in probability), which corresponds to the frailty-less case (the i.i.d. case), we recover the estimator in [3], whereas, if $\beta\to 0$, we recover the estimator in [19]. Notice here also that at the degenerate case, the reduced model still accounts for informative monitoring because of the presence of the term $\beta I\{\tau_i\ge s\}$. With the choice of gamma distribution with equal parameters for the frailties, and from (3), the common marginal survival function is given by

$$\hat{\bar{F}}(s,t|\hat{\beta};\mathbf{z})=\left[\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\Lambda}_0(s,t|\hat{\beta};\mathbf{z})}\right]^{\hat{\alpha}},\tag{9}$$

where $\hat{\alpha}$ is the MLE of α obtained via the procedure described in Section 4. The asymptotic behavior of $\hat{\Lambda}_0(t|\hat{\beta};\mathbf{z})$ can be obtained by extending the results of [13] to recurrent events. We defer this development to future work. For now, we are content to investigate its finite sample behavior through simulation in Section 5.

In the upcoming sections, denote by $\check{\Lambda}(s,t|\check{\beta})$ the estimator derived under informative monitoring in the i.i.d. case, that is with $Z_i\equiv 1$.

2.4. *Inconsistency of* $\check{\Lambda}(s,\cdot|\check{\beta})$.

The product-limit estimator $\check{\Lambda}(s,t|\check{\beta})$ in [3] was developed under the i.i.d. assumption for the recurrence times and is expected to yield a biased estimator of the marginal survivor function for correlated recurrence times. Peña et al. [17] addressed the consequences of analyzing datasets generated from a model with frailty using procedures developed from the model without frailty (under-specification) and vice versa. They have shown, using a simulation study that under-specification could lead to non-negligible systematic biases in the resulting estimators.

In [3], a semiparametric estimator of the cumulative hazard without frailty was given by

$$\check{\Lambda}(s,t|\check{\beta})=\int_0^t J(s,w;\check{\beta})\left\{\frac{N(s,dw)+N^\tau(dw)}{Y(s,w)+\check{\beta}Y^\tau(w)}\right\},\tag{10}$$

where $J(s,t;\check{\beta}) = I\{Y(s,w)+\check{\beta}Y^{\tau}(w) > 0\}$ and $\check{\beta}$ is the MLE under the i.i.d. assumption. The following result characterizes the asymptotic behavior of $\check{\Lambda}(s,t|\check{\beta})$ when the inter-event times satisfy the frailty model. In what follows, for $s\to\infty$, let $\check{\Lambda}(s,t|\check{\beta}) \equiv \check{\Lambda}(\infty,t|\check{\beta}) \equiv \check{\Lambda}(t)$.

Theorem 1. *Let $t^{\star} > 0$ and $t \le t^{\star}$. Assume $E\{(Y(s,t^{\star})+\beta Y^{\tau}(t^{\star}))/n\} > 0$ and $\Lambda(t^{\star}) < \infty$. Then, under the generalized KG model and the frailty model, $\check{\Lambda}(t) \xrightarrow{p} \int_0^t \phi(\infty,w)\lambda_0(w)dw$ as $n\to\infty$, where*

$$\phi(\infty,w) = \frac{E\{Z_1(y(w|\bar{F}_0^{Z_1}) + \beta\bar{F}_0(w)^{\beta Z_1})\}}{E\{y(w|\bar{F}_0^{Z_1}) + \beta\bar{F}_0(w)^{\beta Z_1}\}}.$$

Proof. The expression in (10) can also be written as

$$\check{\Lambda}(s,t|\check{\beta}) = \sum_{i=1}^{n}\int_0^t \frac{J(s,w;\check{\beta})(M_i(s,dw|Z_i)+M_i^{\tau}(dw|Z_i))}{Y(s,w)+\check{\beta}Y^{\tau}(w)} + \sum_{i=1}^{n}\int_0^t \frac{Z_i\{Y_i(s,w)+\check{\beta}Y_i^{\tau}(w)\}}{Y(s,w)+\check{\beta}Y^{\tau}(w)}\lambda_0(w)dw. \quad (11)$$

Proposition 1 of [19] can be used to show that the first term in the right hand side of (11) is a zero-mean square-integrable martingale with respect to $\mathbf{F}$ under the frailty model. Next, consider the second term in the right hand side of (11). We have

$$\sum_{i=1}^{n}\int_0^t \frac{Z_i(Y_i(s,w)+\check{\beta}Y_i^{\tau}(w))}{Y(s,w)+\check{\beta}Y^{\tau}(w)}\lambda_0(w)dw = \sum_{i=1}^{n}\int_0^t \frac{\{Z_i(Y_i(s,w)+\check{\beta}Y_i^{\tau}(w))\}/n}{\{Y(s,w)+\check{\beta}Y^{\tau}(w)\}/n}\lambda_0(w)dw.$$

From the Weak Law of Large Numbers and the consistency of $\check{\beta}$ (cf. [3]), we obtain

$$\sum_{i=1}^{n}\int_0^t \frac{Z_i\{Y_i(s,w)+\check{\beta}Y_i^{\tau}(w)\}/n}{\{Y(s,w)+\check{\beta}Y^{\tau}(w)\}/n}\lambda_0(w)dw \xrightarrow{p} \int_0^t \frac{E\{Z_1(Y_1(s,w)+\beta Y_1^{\tau}(w))\}}{E\{Y_1(s,w)+\beta Y_1^{\tau}(w)\}}\lambda_0(w)dw,$$

Let $s=\infty$ and set $Y(\infty,t)\equiv Y(t)$. Using the iterated expectation, we have $E\{Z_1Y_1(w)\} = E\{Z_1E(Y(w)|Z_1)\} = E\{Z_1 y(w|\bar{F}_0^{Z_1})\}$ where $y(w|\bar{F}_0^{Z_1}) = E(Y(w)|Z_1)$ and $E\{Z_1Y_1^{\tau}(w)\} = E\{Z_1\bar{F}_0(w)^{\beta Z_1}\}$. To find $y(t|\bar{F}_0^{Z_1})$, apply Proposition 2 of [19] to obtain

$$y(t|\bar{F}_0^{Z_1}) = \bar{F}_0^{Z_1}(t)\bar{G}(t-) + \bar{F}(t)\int_t^{\infty}\rho(w-t|\bar{F}_0^{Z_1})dG(w),$$

where $\bar{G}(w) = \bar{F}^{Z_1\beta}(w)$ is the survival function for the monitoring time given Z_1 and $\rho(\cdot|\bar{F}_0^{Z_1})$ is the renewal function associated with the conditional gap time survival function $\bar{F}(\cdot|Z_1) = \bar{F}_0(\cdot)^{Z_1}$. □

2.5. Theoretical variance of $\hat{\Lambda}_0(s,\cdot|\hat{\beta},\mathbf{z})$

Given $\mathbf{z}$, β and for fixed t, the covariation process of the hat estimator $\hat{\Lambda}_0(s,\cdot)|\hat{\beta},\mathbf{z}$ is:

$$\langle\hat{\Lambda}_0(\cdot,t|\hat{\beta},\mathbf{z}),\hat{\Lambda}_0(\cdot,t|\beta,\mathbf{z})\rangle(\cdot) = \int_0^t \frac{J(s,w|\beta,\mathbf{z})}{\sum_{i=1}^{n} z_i\{Y_i(s,w)+\beta Y_i^{\tau}(w)\}}\lambda_0(w)dw.$$

For fixed (s,t), under the gamma frailty model, let $K(\Lambda(s,t),\alpha) = \left\{1+\frac{\Lambda(s,t)}{\alpha}\right\}^{-\alpha}$. We can rewrite the hat estimator as:

$$\hat{\hat{\bar{F}}}(s,t|\alpha,\beta,\mathbf{z}) = K(\Lambda_0(s,t),\alpha) + K'(\Lambda^{*}(s,t),\alpha)\left(\hat{\Lambda}(s,t|\beta,\mathbf{z}) - \Lambda_0(s,t)\right),$$

where $\Lambda^*(s,t) \in \left(\hat{\Lambda}(s,t|\beta,\mathbf{z}), \Lambda_0(s,t)\right)$ and $K'(\Lambda^*(s,t),\alpha)$ is the partial derivative of $K(\cdot)$ with respect to $\Lambda(s,t)$ for fixed (s,t). Hence,

$$V(\hat{\bar{F}}(s,t|\alpha,\beta,\mathbf{z})) = K'(\Lambda^\star(s,t),\alpha)^2 \int_0^t \frac{J(s,w|\beta,\mathbf{z})}{\sum_{i=1}^n z_i\{Y_i(s,w)+\beta Y_i^\tau(w)\}}\lambda_0(w)dw.$$

Consequently, the asymptotic variance of $\sqrt{n}V(\hat{\bar{F}}(s,t|\alpha,\beta,\mathbf{z}))$ becomes

$$V(\sqrt{n}\hat{\bar{F}}(s,t|\alpha,\beta,\mathbf{z})) \xrightarrow{p} K'(\Lambda_0(s,t),\alpha)^2 \int_0^t \frac{P\left(Z_1Y_1(s,w)+Z_1\beta Y_1^\tau(w)>0\right)}{E\{Z_1(y(w|\bar{F}_0^{Z_1})+\beta\bar{F}_0(w)^{\beta Z_1})\}}\lambda_0(w)dw,$$

where

$$y(w|\bar{F}_0^{Z_1}) = \bar{F}_0^{Z_1}(w)\bar{F}_0(w)^{Z_1\beta} + \bar{F}_0^{Z_1}(w)\int_w^\infty \rho(v-w|\bar{F}_0^{Z_1})dG(v).$$

Using similar arguments, the asymptotic variance of the Kaplan–Meier estimator with frailty [19] $\sqrt{n}V(\tilde{\bar{F}}(s,t|\alpha,\beta,\mathbf{z}))$ under the KG model is

$$V(\sqrt{n}\tilde{\bar{F}}(s,t|\alpha,\mathbf{z})) \xrightarrow{p} K'(\Lambda_0(s,t),\alpha)^2 \int_0^t \frac{P\left(Z_1Y_1(s,w)>0\right)}{E\{Z_1 y(w|\bar{F}_0^{Z_1})\}}\lambda_0(w)dw. \quad (12)$$

3. Estimation of frailties and model parameters

As developed in Section 2.3, the nonparametric estimator of the cumulative hazard under informative monitoring and the frailty models is given by

$$\hat{\Lambda}_0(s,t|\hat{\beta};\mathbf{z}) = \int_0^t J(s,w;\hat{\beta})\left\{\frac{N(s,dw)+N^\tau(dw)}{\sum_{i=1}^n z_i\{Y_i(s,w)+\hat{\beta}Y_i^\tau(w)\}}\right\}. \quad (13)$$

Because the Z_is are unobservable, $\hat{\Lambda}_0(s,t|\hat{\beta},z_1,\ldots,z_n)$ cannot be computed. The goal of this section is to obtain an estimate of the unobserved frailties. We begin with the full data conditional likelihood over $[0,s]$. Following [9], we have

$$\begin{aligned} L(s,\beta|\mathbf{z}) = &\prod_{i=1}^n\prod_{w=0}^s \{z_iY_i^\dagger(w)\lambda_0(R_i(w))dw\}^{dN_i^\dagger(w)}\{1-z_iY_i^\dagger(w)\lambda_0(w)dw\}^{1-dN_i^\dagger(w)} \\ &\times \prod_{w=0}^s \{z_i\beta Y_i^\dagger(w)\lambda_0(w)dw\}^{dN_i^\tau(w)}\{1-z_i\beta Y_i^\dagger(w)\lambda_0(w)dw\}^{1-dN_i^\tau(w)}. \end{aligned}$$

Applying Bayes' rule by multiplying the likelihood $L(s,\beta|\mathbf{z})$ by the joint density of $\mathbf{z}=(z_1,\ldots,z_n)$ (each z_i has density $\alpha^\alpha z_i^{\alpha-1}\exp\{-\alpha z_i\}/\Gamma(\alpha)$), we obtain

$$\begin{aligned} L(s,\beta;\mathbf{z}) &= L(s,\beta|\mathbf{z}) \times \prod_{i=1}^n f(z_i) \\ &= \prod_{i=1}^n\prod_{w=0}^s \{z_iY_i^\dagger(w)\lambda_0(R_i(w))dw\}^{dN_i^\dagger(w)} \prod_{w=0}^s \{z_i\beta Y_i^\dagger(w)\lambda_0(w)dw\}^{dN_i^\tau(w)} \\ &\quad\times \exp\left\{-z_i\int_0^s Y_i^\dagger(w)\{\lambda_0(R_i(w))+\beta\lambda_0(w)\}dw\right\} \times \frac{\alpha^\alpha z_i^{\alpha-1}\exp(-\alpha z_i)}{\Gamma(\alpha)}, \end{aligned}$$

which can also be written

$$L(s,\beta;z_1,\ldots,z_n) = \prod_{i=1}^{n} \beta^{N_i^{\tau}(s)} \frac{\alpha^{\alpha}}{\Gamma(\alpha)} z_i^{N_i^{\dagger}(s)+N_i^{\tau}(s)+\alpha-1} \times \prod_{w=0}^{s}\{Y_i^{\dagger}(w)\lambda_0(R_i(w))dw\}^{dN_i^{\dagger}(w)} \prod_{w=0}^{s}\{\beta Y_i^{\dagger}(w)\lambda_0(w)dw\}^{dN_i^{\tau}(w)} \times \exp\left\{-z_i\left\{\alpha+\int_0^s Y_i^{\dagger}(w)\{\lambda_0(R_i(w))+\beta\lambda_0(w)\}dw\right\}\right\}. \quad (14)$$

The likelihood in (14) is proportional to that of a gamma distribution with shape parameter $N_i^{\dagger}(s) + N_i^{\tau}(s) + \alpha$ and scale parameter $\alpha + \int_0^s Y_i^{\dagger}(w)\{\lambda_0(R_i(w)) + \beta\lambda_0(w)\}dw$. As a consequence, if we let $\mathbf{P}_i = \{Y_i^{\dagger}(w), N_i^{\dagger}(w), N_i^{\tau}(w);\ w \in [0,s]\}$, $i = 1,\ldots,n$, then

$$E\{Z_i|\alpha,\beta,\Lambda_0(\cdot),\mathbf{P}_i\} = \frac{N_i^{\dagger}(s)+N_i^{\tau}(s)+\alpha}{\alpha+\int_0^s Y_i^{\dagger}(w)\{\lambda_0(R_i(w))+\beta\lambda_0(w)\}dw}. \quad (15)$$

To obtain an estimate of the association parameter α and the monitoring parameter β one can integrate out $z_1,\ldots,z_n$ in $L(s,\beta;\mathbf{z})$ and the corresponding marginal likelihood function is

$$L(s;\beta,\alpha) = \prod_{i=1}^{n} \frac{\Gamma(N_i^{\dagger}(s)+N_i^{\tau}(s)+\alpha)}{\Gamma(\alpha)} \times \left\{\frac{\alpha}{\alpha+\int_0^s Y_i^{\dagger}(w)\{\lambda_0(R_i(w))+\beta\lambda_0(w)\}dw}\right\}^{N_i^{\dagger}(s)+N_i^{\tau}(s)+\alpha} \times \prod_{w=0}^{s}\left\{\frac{Y_i^{\dagger}(w)\lambda_0(R_i(w))}{\alpha}\right\}^{dN_i^{\dagger}(w)} \prod_{w=0}^{s}\left\{\frac{Y_i^{\dagger}(w)\lambda_0(w)}{\alpha}\right\}^{dN_i^{\tau}(w)} \beta^{N_i^{\tau}(s)}. \quad (16)$$

$L(s;\beta,\alpha)$ can be maximized with respect to α, β, and $\lambda_0(\cdot)$ by mimicking the expectation and maximization (EM) algorithm for missing data problems as done in [15] for gamma frailty models. The next section details this application of the EM algorithm.

4. Computational forms

In this section, we will show how our estimator is implemented using any recurrent event data. Let $\mathbf{U} = \{u_j\}$ be a $Q \times 1$ vector of pooled gap times and monitoring period given by,

$$\mathbf{U}^t = \bigcup_{i=1}^{n}\left\{\tau_i, T_{i1}, T_{i2}, \ldots, T_{iK_i}\right\}.$$

Furthermore, define the $Q \times n$ at-risk matrices by $\mathbf{Y}^{\dagger}(\mathbf{U}) = \left\{y_{ij}^{\dagger}\right\}$ and $\mathbf{Y}(s,\mathbf{U}) = \left\{y_{ij}(s)\right\}$, where $y_{ij}^{\dagger} = Y_i^{\dagger}(u_j)$ is defined in (4) and $y_{ij}(s) = Y_i(s,u_j)$ is defined in (7). Define the $Q \times 1$ jump vectors $\mathbf{d}(s) = \left\{d_j(s)\right\} = \left\{N(s,\Delta u_j)\right\}$ and $\mathbf{d}^{\tau} = \left\{d_j^{\tau}\right\} = \left\{N^{\tau}(\Delta u_j)\right\}$, where

$$d_j(s) = \sum_{i=1}^{n}\sum_{m=1}^{N_i^{\dagger}(s)} I\left(T_{im}=u_j\right) \quad \text{and} \quad d_j^{\tau} = \sum_{i=1}^{n} I\left(\tau_i = u_j\right).$$

Given an $n \times 1$ frailty vector $\mathbf{Z}$ and the parameter β, the computational form of the cumulative hazard in (13) for $t \in \mathbf{U}$ is $\hat{\Lambda}_0(s,t|\beta,\mathbf{z}) = \sum_{j|u_j\le t}\hat{\lambda}_0(s,u_j|\beta,\mathbf{z})$, where

$$\hat{\lambda}_0(s,u_j|\beta,\mathbf{z}) = \frac{d_j(s)+d_j^{\tau}}{\sum_{i=1}^{n} z_i\left(y_{ij}(s)+\beta y_{ij}^{\dagger}\right)}.$$

The reason for estimating the cumulative hazard using this approach is that the parametric estimation was not amenable to any closed-form mathematical estimation. The modified EM algorithm proposed by Peña et al. [19] is implemented in the following way to obtain $(\hat{\alpha}, \hat{\beta})$. Start with the initial values:

$$\alpha^{(0)} = 1, \quad \beta^{(0)} = 0.001, \quad \text{and} \quad \mathbf{Z}^{(0)} = \{Z_i = 1 : i = 1, 2, \ldots, n\}.$$

Consider the whole recurrent data set by letting $s = s^* = \max(\mathbf{U})$. The E-step involves generating the $n \times 1$ vector of posterior mean frailties given previous values of (α, β); that is, $\mathbf{Z}^{(q+1)} = \left\{Z_i^{(q+1)}\right\}, i = 1, \ldots, n$, where

$$Z_i^{(q+1)} = \frac{K_i + 1 + \alpha^{(q)}}{\alpha^{(q)} + \sum_{j=1}^{Q} \left(y_{ij}(s*) + \beta^{(q)} y_{ij}^{\dagger}\right) \hat{\lambda}_0(s*, u_j | \beta^{(q)}, \mathbf{Z}^{(q)})}.$$

In the M-step, maximum likelihood estimates of $(\alpha^{(q+1)}, \beta^{(q+1)})$ are obtained satisfying the condition

$$\left(\hat{\alpha}^{(q+1)}, \hat{\beta}^{(q+1)}\right) = \arg\max_{(\alpha,\beta)} \left\{ l_F \left(\alpha, \beta | \hat{\lambda}_0(\cdot), \mathbf{Z}^{(q+1)}\right)\right\},$$

where

$$\begin{aligned} l_F \left(\alpha, \beta | \hat{\lambda}_0(\cdot), \mathbf{Z}^{(q+1)}\right) = & \sum_{i=1}^{n} \sum_{j=0}^{K_i} \log(j + \alpha) + n\alpha \log(\alpha) \\ & + \sum_{j=1}^{Q} \log\left(\widehat{\lambda}\left(s, u_j | \beta, \mathbf{Z}^{(q+1)}\right)\right) \left[d_j + d_j^{\tau}\right] \\ & - \sum_{i=1}^{n} (K_i + 1 + \alpha) \log\left(\alpha + \sum_{j=1}^{Q} \widehat{\lambda}\left(s, u_j | \beta, \mathbf{Z}^{(q+1)}\right)\left[y_{ij}(s*) + \beta y_{ij}^{\dagger}\right]\right) \end{aligned} \tag{17}$$

is the log-profile likelihood in (16) with $\hat{\lambda}_0(\cdot)$ in place of $\lambda_0(\cdot)$. The solution to the maximization of L_F may be obtained using Newton–Raphson or Nelder–Mead simplex methods [14].

5. Simulation studies

Simulation studies were performed to examine small to moderate sample size properties of two survival estimators: the nonparametric estimator with frailty denoted by $\tilde{\tilde{F}}(s, t)$ proposed in [19], and the semiparametric survival estimator denoted by $\hat{\hat{F}}(s, t)$, which accounts for frailty and informative monitoring. They are given by

$$\tilde{\tilde{F}}(s, t) = \left\{1 + \frac{\tilde{\Lambda}_0(s, t|\tilde{\mathbf{z}})}{\tilde{\alpha}}\right\}^{-\tilde{\alpha}} \quad \text{and} \quad \hat{\hat{F}}(s, t) = \left\{1 + \frac{\hat{\Lambda}_0(s, t|\hat{\beta}, \hat{\mathbf{z}})}{\hat{\alpha}}\right\}^{-\hat{\alpha}}, \tag{18}$$

respectively. Here $\tilde{\Lambda}_0(s, t|\tilde{\mathbf{z}}) = \sum_{j|u_j \le t} \tilde{\lambda}(s, u_j|\tilde{\mathbf{z}})$ such that

$$\tilde{\lambda}(s, u_j|\tilde{\mathbf{z}}) = \frac{d_j(s)}{\sum_{i=1}^{n} y_{ij}(s)\tilde{z}_i},$$

where $\{\tilde{\alpha}, \tilde{\mathbf{z}}\}$ and $\{(\hat{\alpha}, \hat{\beta}), \hat{\mathbf{Z}}\}$ are the ML and frailties estimates at convergence of the EM algorithm. The hat and tilde survival estimators were compared by generating recurrent event data

$$\mathbf{O}_{qr} = \left\{\left(K_i, \tau_i, T_{i1}, T_{i2}, \ldots, T_{iK_i}, \tau_i - S_{iK_i}\right) : i = 1, 2, \ldots, n_q\right\}_r,$$

where q is one of the 54 configurations described below:

$i = 1, 2, \ldots, n$ for $n \in \{20, 50\}$;
$z_i \sim \text{Gamma}(\text{shape} = \alpha, \text{scale} = \alpha)$ for $\alpha \in \{2, 6, \infty\}$;
$T_{ij} \sim F(t|z_i)$ such that $\bar{F}(t|z_i) = \bar{F}(t)^{z_i}$ and
$F(t) \sim \text{Weibull}(\text{shape} = \eta, \text{scale} = 1)$ for $\eta \in \{0.9, 1, 2\}$; and
$\tau_i \sim G(t|\beta, z_i)$ such that $\bar{G}(t|\beta, z_i) = \bar{F}(t)^{z_i\beta}$ for $\beta \in \{0.3, 0.5, 0.7\}$,

and $r = 1, 2, \ldots, M$, $(M = 2000)$ is the replication. The simulated data is constructed under the informative monitoring model with and without frailty. For each replicate, the hat and tilde survival estimates are calculated for time points $\{t_{0.02}, t_{0.04}, \ldots, t_{0.98}\}$ where t_p is the pth percentile of F. Thus, our empirical mean square errors (MSE) with respect to our estimators are indexed by the qth configuration and pth percentile:

$$\widehat{\text{MSE}}_q(\bar{F}(\cdot), \overset{*}{\bar{F}}(\cdot), t_p) = \frac{\sum_{r=1}^{M} \{\overset{*}{\bar{F}}_{qr}(t_p) - (1-p)\}^2}{M} = \frac{\sum_{r=1}^{M} \{\overset{*}{\bar{F}}_{qr}(t_p) - \overset{*}{\mu}_q(t_p)\}^2}{M} + \{\overset{*}{\mu}_q(t_p) - (1-p)\}^2, \quad (19)$$

where $\overset{*}{\bar{F}}(\cdot)$ can be a hat or a tilde estimator such that $\overset{*}{\bar{F}}_{qr}(\cdot)$ is the hat or tilde estimator for the data set $\mathbf{O}_{qr}$, and $\overset{*}{\mu}_q(t_p) = M^{-1} \sum_{r=1}^{M} \overset{*}{\bar{F}}_{qr}(t_p)$. Observe that the first and second terms of (19) are the empirical variance and squared bias, respectively. Efficiency comparisons are represented using empirical log-efficiency of the tilde estimator over the hat estimator:

$$\widehat{\text{eff}}_q\left(\widehat{\bar{F}}(\cdot), \widetilde{\bar{F}}(\cdot), t_p\right) = \log\left\{\widehat{\text{MSE}}_q(\bar{F}(\cdot), \widehat{\bar{F}}(\cdot), t_p)\right\} - \log\left\{\widehat{\text{MSE}}_q(\bar{F}(\cdot), \widetilde{\bar{F}}(\cdot), t_p)\right\},$$

where positive values imply the tilde estimators are more efficient over the hat estimators and negative values implying otherwise.

For brevity, 36 out of the 54 configurations are represented by omitting $\alpha = 1$ in Fig. 1. The graphs are laid out in a grid such that the first two columns pertain to survival function biases for $n = 20$ and $n = 50$, respectively, and the last column contains the empirical log-efficiencies. The horizontal axis is the p index which is equivalent to the true percentile values $\{F(t) : F(t) = 0.02, 0.04, \ldots, 0.98\}$. Results in Fig. 1(a) and (b) suggest that the tilde estimator is more negatively biased than the hat estimator for all configurations in $t \in (t_{0.05}, t_{0.60})$ resulting in less efficiency over the hat estimator. The opposite trend is observed for $t > t_{.6}$ where the hat estimates have larger positive bias than the tilde estimates. From a theoretical viewpoint, this positive bias for the generalized KG estimator is brought about by $\phi(w)$ which is decreasing for increasing w resulting in under-estimation of the cumulative hazard, thereby over-estimating the survivor function. On the other hand, the tilde estimator is performing much better with bias values closer to zero and relative efficiency values greater than one. Since the asymptotic efficiency is

$$\text{eff}_q\left(\widehat{\bar{F}}(s,\cdot), \widetilde{\bar{F}}(s,\cdot), t_p\right) \xrightarrow{p} \log\left\{\int_0^t \frac{P\left(Z_1 Y_1(s,w) + Z_1\beta Y_1^{\tau}(w) > 0\right)}{E\{Z_1(y(w|\bar{F}_0^{Z_1}) + \beta\bar{F}_0(w)^{\beta Z_1})\}} \lambda_0(w)dw\right\} - \log\left\{\int_0^t \frac{P\left(Z_1 Y_1(s,w) > 0\right)}{E\{Z_1(y(w|\bar{F}_0^{Z_1}))\}} \lambda_0(w)dw\right\},$$

a negative log-efficiency on the lower t-values is brought about by

$$E\{Z_1(y(w|\bar{F}_0^{Z_1}) + \beta\bar{F}_0(w)^{\beta Z_1})\} \geq E\{Z_1(y(w|\bar{F}_0^{Z_1}))\}$$

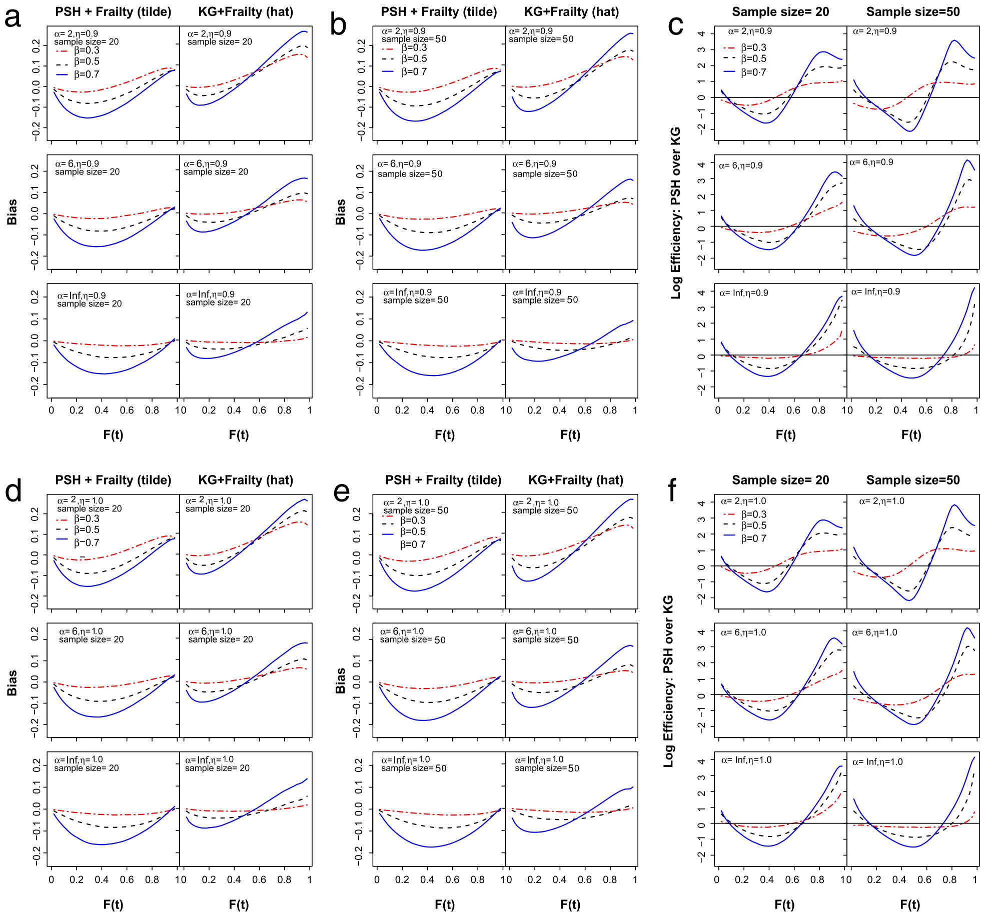

Fig. 1. Simulation results for small and moderate sample sizes. Figures (a) and (b) are the biases for the hat and tilde estimators under the Weibull(0.9,1) for $n = 20$ and $n = 50$, respectively, and (c) is the log-efficiency under Weibull(0.9,1) for both sample sizes. Similarly, Figures (d), (e) and (f) are the two biases under $n = 20$ and 50 and log-efficiency respectively under Weibull(1,1) which is the unit exponential model.

and the positive log-efficiency on the tails is brought about by

$$P\left(Z_1 Y_1(s, w) + Z_1 \beta Y_1^{\tau}(w) > 0\right) \geq P\left(Z_1 Y_1(s, w) > 0\right)$$

The practical implication of the simulation results is that under the generalized KG model, the hat estimator was able to perform well for shorter inter-occurrence times, but it overestimates on longer ones. It is therefore desirable to use a combined estimator, say $\dot{F}(t) = p\hat{F}(t) + (1 - p)\tilde{F}(t)$, where $p = 1$ if both $\hat{F}(t)$ and $\tilde{F}(t)$ are both less than or equal to .6; and use $p = 0$ if both are greater than $p = .6$. Use $p = .5$ if one of the estimates is greater than .6 and the other is less than .6. Here $p \in (0, 1)$ is either predetermined or data-driven.

6. Application

The Migrating Motor Complex (MMC) [2] gastroenterology data is used to demonstrate the applicability in a real data setting. This data contains observations where 19 individuals are observed

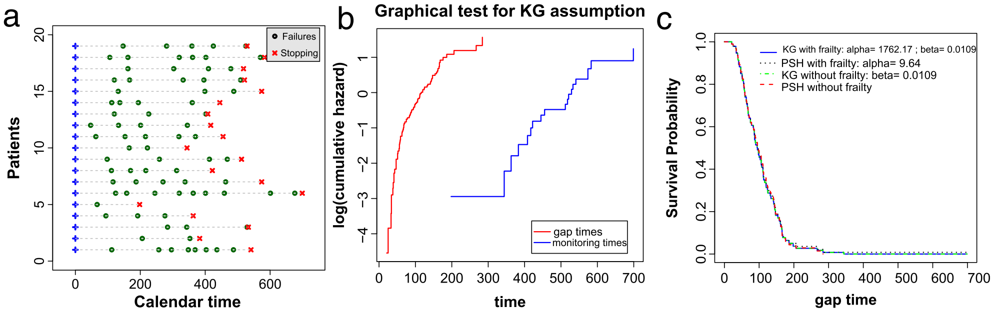

Fig. 2. (a) A pictorial representation of the Migratory Motor Complex, (b) proportionality hazards plot for the gap and monitoring times, and (c) the survival estimates.

during a fasting state. Calendar times are in minutes and gap times represent the time to the next observed event (Fig. 2(a)).

A thorough mathematical development for testing the KG assumption is a daunting task and is currently under investigation. For the time being, we have chosen a heuristic and graphical approach to check this assumption. The validity of the model may be checked by plotting a non-parametric estimate of the log cumulative hazards separately for inter-event times and monitoring times, and using the pictorial representation to detect departure from the generalized KG model. In case differences between these two hazards are constant over the domain where both curves overlap, we will find graphical evidence in favor of the generalized KG model. To check the assumption with the MMC data, a plot of the natural log of the non-parametric cumulative hazard [19] was created separately for gap times and monitoring times and this graph was used to detect departures from the proportional hazard assumption (Fig. 2(b)). On the interval where the curves overlap, the differences are fairly stable; hence, we do not find any serious evidence against our model.

A plot of the semiparametric survival based on informative monitoring with frailty was used for comparison with the non-parametric survival without informative monitoring and without frailty [19], the non-parametric survival without informative monitoring but with frailty [19], and semiparametric survival with informative monitoring but without frailty [3] in Fig. 2c. Our estimator yielded $\hat{\alpha} = 1752.17$ as compared to [19]'s estimator under frailty of $\tilde{\alpha} = 9.64$, which was slightly different from the published value of $\tilde{\alpha} = 10.17$. The high α estimate resulted in a very small variance estimate of 0.0005 implying that the estimated frailties do not differ significantly from 1. As to the parameter estimate, our $\hat{\beta}$ was 0.0109 which was the same as the estimated β for the informative monitoring without frailty. The small value of β means there are a high number of gap times within the monitoring times. Because of the high α and low β, all the curves are expected to be close as shown in Fig. 2(c).

7. Concluding remarks

In this manuscript, we presented a recurrent event framework such that the gap times and the monitoring times are reconciled through the generalized KG model and using frailties to account for subject-specific variability. Using a stochastic process formulation, we obtained the semiparametric survival estimator and its asymptotic variance. Similar to previous findings by Peña et al. [19], the semiparametric estimator in [3] is inconsistent when the recurrence times are correlated but can be corrected by subtracting the bias component as shown in Theorem 1. Simulation results comparing the uncorrected (hat) estimator and the semiparametric (tilde) estimator without frailty for small and moderate samples under Weibull gap times with frailty and with informative monitoring showed that either one can be more efficient than the other which is consistent with the asymptotic efficiency result. Finally, we used the Migratory Motor Complex data and compared our estimator to other estimators developed under a no-frailty or without informative monitoring. In the MMC data, all survival curves were close due to long monitoring periods in comparison to gap times and small variation in our estimated frailties.

Acknowledgments

The authors are grateful to the editor in chief and the referees for a number of suggestions for improvement. The authors thank Dr. John Singler for his careful reading of the manuscript. Dr. Akim Adekpedjou acknowledges research support from the Missouri Research Board (MRB) grant No. DEPTID R 6016058 MOCODE RC460 and Dr. Jonathan Quiton acknowledges research support from the Kentucky EPSCoR Research Startup Fund No. RSF-031-06.

References

[1] O. Aalen, Nonparametric inference for a family of counting processes, Ann. Statist. 6 (1978) 701–726.
[2] O. Aalen, E. Husbeye, Statistical analysis of repeated event forming renewal processes, Stat. Med. 10 (1991) 1227–1240.
[3] A. Adekpedjou, Estimation with recurrent event data under informative monitoring, Ph.D. Thesis, University of South Carolina, 2007.
[4] A. Adekpedjou, E. Peña, J. Quiton, Estimation and efficiency with recurrent event data under informative monitoring, J. Statist. Plann. Inference 140 (2010) 597–615.
[5] P.E. Cheng, G.D. Lin, Maximum likelihood estimation of a survival function under the Koziol–Green proportional hazards model, Statist. Probab. Lett. 5 (1987) 75–80.
[6] R.D. Gill, Nonparametric estimation based on censored observations of a Markov renewal process, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 53 (1980) 97–116.
[7] R.D. Gill, Testing with replacement and the product limit estimator, Ann. Statist. 9 (1981) 853–860.
[8] C.-Y. Huang, M.-C. Wang, Joint modeling and estimation for recurrent event processes and failure time data, J. Amer. Statist. Assoc. 99 (2004) 1153–1165.
[9] J. Jacod, Multivariate point processes: predictable projection, Radon-Nikodým derivatives, representation of martingales, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 31 (1974–75) 235–253.
[10] E.L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, J. Amer. Statist. Assoc. 53 (1958) 457–481.
[11] J.A. Koziol, S.B. Green, A Cramér-von Mises statistic for randomly censored data, Biometrika 63 (1976) 465–474.
[12] L. Liu, R.A. Wolfe, X. Huang, Shared frailty models for recurrent events and a terminal event, Biometrics 60 (2004) 747–756.
[13] S.A. Murphy, Asymptotic theory for the frailty model, Ann. Statist. 23 (1995) 182–198.
[14] J. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 308–313.
[15] G. Nielsen, R. Gill, P. Andersen, T. Sorensen, A counting process approach to maximum likelihood estimation in frailty models, Scand. J. Stat. 19 (1992) 25–43.
[16] E.A. Peña, M. Hollander, Models for recurrent events in reliability and survival analysis, in: Mathematical Reliability: An Expository Perspective, in: Internat. Ser. Oper. Res. Management Sci., vol. 67, Kluwer Ad. Publ., Boston, MA, 2004, pp. 105–123. Chap.
[17] E.A. Peña, E.H. Slate, J.R. Gonzalez, Semiparametric inference for a general class of models for recurrent events, J. Statist. Plann. Inference 137 (2007) 1727–1747.
[18] E.A. Peña, R.L. Strawderman, M. Hollander, A weak convergence result relevant in recurrent and renewal models, in: Recent Advances in Reliability Theory (Bordeaux, 2000), Birkhäuser Boston, Stat. Ind. Technol, Boston, MA, 2000, pp. 493–514.
[19] E.A. Peña, R.L. Strawderman, M. Hollander, Nonparametric estimation with recurrent event data, J. Amer. Statist. Assoc. 96 (2001) 1299–1315.
[20] T. Sellke, Weak convergence of the Aalen estimator for a censored renewal process, in: Statistical Decision Theory and Related Topics, IV, Vol. 2 (West Lafayette, Ind., 1986), Springer, New York, 1988, pp. 183–194.
[21] v.d. Vaart, R. Gill, Non and semiparametric maximum likelihood estimators and the von Mises method-II, Scand. J. Stat. 20 (1993) 271–288.
[22] Y. Vardi, Nonparametric estimation in renewal processes, Ann. Statist. 10 (1982) 772–785.
[23] Y. Vardi, Nonparametric estimation in the presence of length bias, Ann. Statist. 10 (1982) 616–620.
[24] M.-C. Wang, S.-H. Chang, Nonparametric estimation of a recurrent survival function, J. Amer. Statist. Assoc. 94 (1999) 146–153.
[25] K.D. Zamba, A. Adekpedjou, Parameter estimation with correlated recurrent event data under informative monitoring, Statistical Methodology 8 (2011) 273–290.