JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Johns Hopkins University, Dept. of Biostatistics Working Papers

11-1-2003

Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data

Chiung-Yu Huang
Division of Biostatistics, School of Public Health, University of Minnesota, cyhuang@biostat.umn.edu

Mei-Cheng Wang
Johns Hopkins Bloomberg School of Public Health, mcwang@jhsph.edu

Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data

Chiung-Yu Huang and Mei-Cheng Wang

Abstract

Recurrent event data are commonly encountered in longitudinal follow-up studies related to biomedical science, econometrics, reliability, and demography. In many studies, recurrent events serve as important measurements for evaluating disease progression, health deterioration, or insurance risk. When analyzing recurrent event data, an independent censoring condition is typically required for the construction of statistical methods. Nevertheless, in some situations, the terminating time for observing recurrent events could be correlated with the recurrent event process and, as a result, the assumption of independent censoring is violated. In this paper, we consider joint modeling of a recurrent event process and a failure time in which a common subject-specific latent variable is used to model the association between the intensity of the recurrent event process and the hazard of the failure time. The proposed joint model is flexible in that no parametric assumptions on the distributions of censoring times and latent variables are made and, under the model, informative censoring is allowed for observing both the recurrent events and failure times. We propose a 'borrow-strength estimation procedure' by first estimating the value of the latent variable from recurrent event data, and next using the estimated value in the failure time model. Some interesting implications and trajectories of the proposed model will be presented. Properties of the regression parameter estimates and the estimated baseline cumulative hazard functions are also studied.

KEY WORDS: Borrow-strength method; Frailty; Informative censoring; Joint model; Non-stationary Poisson process;.

1

1 INTRODUCTION

Recurrent event data are often collected in longitudinal follow-up studies. During the observation period, recurrent events such as repeated tumor occurrences (Byar 1980), repeated hospitalizations (Eaton et al. 1992a,b), or recurrent injuries (Wassel et al. 1999) are recorded in the studies. The observation of recurrent events could be terminated (i.e., censored) by loss to follow-up, end of the study, or a failure event such as death. Conventional analysis usually focuses on either failure time data (Cox 1972; Cox and Oakes 1984) or recurrent event data (Prentice, Williams, and Peterson 1981; Andersen and Gill 1982; Pepe and Cai 1993; Lin et al. 2000; Wang, Qin, and Chiang 2001). In this paper, the event process and the failure time are both of interest, and we consider the joint modeling of a recurrent event process and a failure time.

In analyzing recurrent event data an independent censoring condition is usually required for the development of statistical methods under different types of models. When a failure event serves as a part of the censoring mechanism, validity of the independent censoring assumption is violated when the recurrent event process is correlated with the failure time. Lancaster and Intrator (1998) considered a joint parametric model of the recurrent event process and the failure time, and demonstrated the use of their methodology using AIDS panel data. In their work, a latent variable is used to characterize the association between the recurrent event process and the failure time, and a common baseline function is shared by the intensity of the recurrent event process and the hazard of the failure time. In nonparametric and semiparametric settings, Wang, Qin, and Chiang (2001) proposed estimation procedures for estimating the cumulative rate function and regression parameters under multiplicative intensity models with dependent censoring. In the WQC model, the focus was placed on the distributional pattern of the recurrent event process where the censoring time was treated as a nuisance, and the joint modeling of recurrent event process and failure time was not

2

considered.

To jointly model recurrent events and failure time, Ghosh and Lin (2003) studied correlated marginal models for these two outcomes. At the cost of censoring some of the originally uncensored data, they developed estimation inferences with the correlation between recurrent events and failure time unspecified. Using a general censoring pattern, Huang and Wang (2003) proposed statistical methods to study two nested joint models of a recurrent event process and a failure time, where the correlation of the two outcomes is partially specified in the conditional distribution of the recurrent event process given the failure time. Note that neither of the two papers used frailty in their joint models. In this paper we consider joint modeling of the recurrent event process and the failure time via frailty. This joint model possesses attractive features of frailty models, especially in its interpretation of correlation, and avoids parametric assumption on the frailty term.

This paper is organized as follows: In Section 2, we introduce a joint model of recurrent event process and failure time in which a common subject-specific latent variable (frailty) is used to model the association between intensity of the recurrent event process and hazard of the failure time. The proposed joint model is flexible in that no parametric assumptions on the distributions of censoring times and latent variables are made and, under the model, informative censoring is allowed for observing both the recurrent events and failure times. Section 3 presents theoretical implications and trajectories of the proposed model. In Section 4, we study a 'borrow-strength estimation procedure' by first estimating the value of the latent variable from recurrent event data, and next using the estimated values in the failure time models. Properties of the regression parameter estimators and the estimated baseline cumulative hazard functions are studied. Section 5 reports results of simulation studies, along with the application to a Denmark schizophrenia case cohort study. A discussion is given in Section 6.

3

2   Notation and the Joint Model

Let $N(t)$ denote the number of events occurring before or at time t, and let D be the failure time and C be the potential censoring time for reasons other than the failure event, respectively. The research interest is to derive inferential results on $N(\cdot)$ and D within a fixed time interval $[0, T_0]$, where the event process could potentially be observed beyond T_0. Let X be a $1 \times p$ vector of covariates. We then make the following model assumptions:

(M1) There exists a non-negative valued latent variable Z so that, given $X = x$ and $Z = z$, the recurrent event process $N(\cdot)$ is a non-stationary Poisson process with intensity function

$$\lambda(t) = z\lambda_0(t) \exp(x\alpha), \qquad 0 \le t \le T_0,$$

where α is a $p \times 1$ vector of parameters and the baseline intensity function $\lambda_0(t)$ is a continuous function with $\Lambda_0(T_0) = \int_0^{T_0} \lambda_0(u)du = 1$. The latent variable Z satisfies $E(Z \mid X) = E(Z)$.

(M2) Given (x, z), the hazard function of D takes the form

$$h(t) = zh_0(t) \exp(x\beta),$$

where β is a $p \times 1$ vector of parameters and the baseline hazard function $h_0(t)$ is continuous.

(M3) Conditioning on (x, z), $(N(\cdot), D, C)$ are mutually independent.

The occurrence of recurrent events is modeled by a subject-specific Poisson process via a latent variable. Conditioning on z, the rate function equals the intensity function since a Poisson process is memoryless. Under (M1), the baseline intensity function $\lambda_0(t)$ is shared by all subjects and is left unspecified. A multiplicative hazard function with the same latent

4

variable but a different baseline function is assumed for the hazard of failure event in (M2). Clearly, large value of z inflates both the intensity of recurrent events and the hazard of the failure event. Under Assumption (M3), D, C and $N(\cdot)$ are allowed to be correlated via their connection with (x, z). This model relaxes the requirement that a common baseline function be shared by the intensity of $N(\cdot)$ and the hazard of D assumed by Lancaster and Intrator (1998), and still keeps the semiparametric model features of Wang et al. (2001). Define $Y = \min(C, D, T_0)$, the time when the observation of the recurrent event process is terminated and $\Delta_i = I(D_i \leq Y_i)$, the observed censoring indicator. By further conditioning on z, the usual independent censoring condition that $N(\cdot)$ be independent of Y given x is relaxed for recurrent events, and, interestingly, the independent censoring condition that D be independent of C given x is also relaxed for failure time data.

Note that the rate function of event occurrence at time t in a random population, for study subjects with explanatory variable x, is $\mu_Z \lambda_0(t) \exp(x\alpha)$, where $\mu_Z = E[Z]$. In many public health and biomedical studies, the rate function is preferred for analysis, especially in identifying treatment effects and risk factors, because of its marginal interpretation. For instance, the parameter α can be interpreted as the logarithm of the ratio of the rate function for every unit increase in the explanatory variable.

Under (M1$\sim$3), the distribution of Z, the baseline functions $\lambda_0(t)$ and $h_0(t)$, and the distribution of C serve as nonparametric components in the model. In the next section we will examine model implications with or without additional parametric assumptions on Z, yet no parametric assumptions will be made on these components for our developments of estimation inferences in Section 4.

5
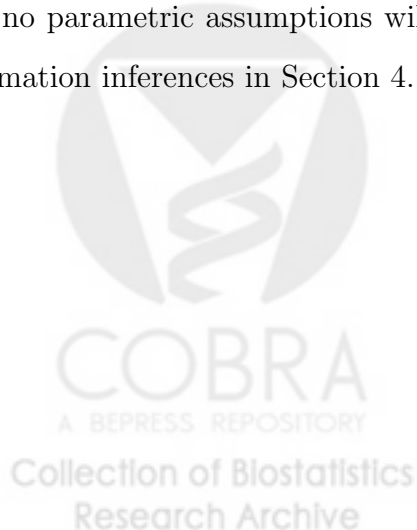
3 MODEL IMPLICATIONS

Let $\mathcal{H}(t) = \{N(u) : 0 \leq u \leq t\}$ be the event history up to t, and let $t_1 \leq t_2 \leq \ldots \leq t_{N(t)}$ be the ordered event times before or at t. Define $f_Z(\cdot)$ to be the probability density function of the latent variable Z, $f(\cdot)$ a general probability density function, and $f(\cdot \mid \cdot)$ a general conditional probability density function. In this section, we will discuss model implications under the proposed joint model with or without additional parametric assumptions on Z. To simplify the discussion, we consider the reduced model without covariates. Similar results for regression models with covariates can be obtained with replacement of $(\Lambda_0(t), \lambda_0(t), H_0(t), h_0(t))$ by $(\Lambda_0(t)e^{x\alpha}, \lambda_0(t)e^{x\alpha}, H_0(t)e^{x\beta}, h_0(t)e^{x\beta})$.

Implication 1. *(Posterior mean of Z)* Given the observed recurrent event data, we show in Appendix A that the posterior mean of Z can be expressed as

$$E[Z \mid \mathcal{H}(y), y] \;=\; \frac{N(y) + 1}{\Lambda_0(y)} \times \frac{f(N(y) + 1 \mid y)}{f(N(y) \mid y)} \;,$$

where $f(N(y) \mid y)$ is the conditional probability density function of $N(Y)$ given Y. One can examine that, given the follow-up time y, the posterior mean depends on the event history $\mathcal{H}(y)$ only through the number of observed events. The posterior mean can be used for individual-specific prediction when additional model assumptions are available for obtaining an explicit form of the formula.

Implication 2. *(Residual life time)* Let t and s be non-negative constants. For individuals who survive beyond time t ($D \geq t$), the conditional probability for the residual life time to be longer than s units of time given $\mathcal{H}(t)$ is

$$P(D \geq t + s \mid \mathcal{H}(t), D \geq t) \;=\; \frac{E\big[e^{-Z\{H_0(t+s) + \Lambda_0(t)\}} Z^{N(t)}\big]}{E\big[e^{-Z\{H_0(t) + \Lambda_0(t)\}} Z^{N(t)}\big]}.$$

The derivation is given in Appendix A. The computation implies that

$$P(D \geq t + s \mid N(t), D \geq t) = P(D \geq t + s \mid \mathcal{H}(t), D \geq t);$$

6

that is, the residual life time probability depends on the event history only through the number of events occurring up to time t. Further, the median residual life time after time t can be obtained by solving $P(D \geq t + s \mid N(t), D \geq t) = 1/2$. The residual life time unconditional on the event history has the survival function

$$P(D \geq t + s \mid D \geq t) = \frac{E[e^{-ZH_0(t+s)}]}{E[e^{-ZH_0(t)}]}.$$

For the specific case that Z is distributed as $gamma(a, b)$ with mean a/b, the residual life time probability, given the event history, has the survivor function

$$P(D \geq t + s \mid N(t), D \geq t) = \Big(\frac{b + H_0(t) + \Lambda_0(t)}{b + H_0(t + s) + \Lambda_0(t)}\Big)^{N(t)+a}.$$

This conditional survival function has the following interesting interpretation: With each additional event occurrence in the time interval $[0, t]$, the survival probability at time $t + s$ is decreased by the constant factor, $\{b + H_0(t) + \Lambda_0(t)\}/\{b + H_0(t + s) + \Lambda_0(t)\}$, where the constant factor has a value between 0 and 1 and it depends on $(H_0(t), H_0(t + s), \Lambda_0(t), b)$. Additionally, with the assumption that Z is distributed as $gamma(a, b)$, the survival function for the residual life time unconditional on the event history can be expressed as

$$P(D \geq t + s \mid D \geq t) = \Big(\frac{b + H_0(t)}{b + H_0(t + s)}\Big)^{a}.$$

It is then interesting to see that $P(D \geq t + s \mid N(t) = 0, D \geq t) \geq P(D \geq t + s \mid D \geq t)$, where the inequality relationship becomes strict if $\Lambda_0(t) > 0$ and $H_0(t + s) > H_0(t)$ as $s > 0$. That is, survivors at time t who experienced no events before t would have higher probability to live s units of residual life time than those population survivors at time t.

Implication 3. *(Residual life time for censored subjects)* It is also possible to examine the residual life time of those who are *censored* at time t given the event history

$$P(D \geq t + s \mid Y = t, \Delta = 0, \mathcal{H}(t)) = \frac{E[e^{-Z\{H_0(t+s)+\Lambda_0(t)\}} Z^{N(t)} f_c(t \mid Z)]}{E[e^{-Z\{H_0(t)+\Lambda_0(t)\}} Z^{N(t)} f_c(t \mid Z)]},$$

where $\Delta = I(D \leq C)$ is the censoring indicator and $f_c(t \mid z)$ is the conditional probability density function of the censoring time, C, given Z. If we assume that the hazard function of C given Z is $\lambda_c(t \mid z) = zg_0(t)$ and that Z is distributed as $gamma(a,b)$, then we have

$$P(D \geq t + s \mid N(t), Y = t, \Delta = 0) = \left(\frac{b + H_0(t) + \Lambda_0(t) + G_0(t)}{b + H_0(t+s) + \Lambda_0(t) + G_0(t)} \right)^{a+N(t)+1},$$

where G_0 is the cumulative distribution function of g_0. With each additional event, the probability of surviving extra s unit of time after being censored at t is decreased by a constant factor, where the constant factor depends on $(H_0(t), H_0(t+s), \Lambda_0(t), G_0(t), b)$.

Implication 4. *(Effect of failure time on recurrent events)* We derive in Appendix A the mean function of the recurrent event process conditional on the failure time. For $t \geq s$,

$$E[N(s) \mid D \geq t] = \frac{E[Ze^{-ZH_0(t)}]}{E[e^{-ZH_0(t)}]} \Lambda_0(s).$$

The mean function given failure time can be decomposed into two parts: one part depends on the baseline cumulative rate function, and the other part depends on the baseline cumulative hazard function and the frailty distribution. The function $E[N(s) \mid D \geq t]$ can be further shown to be decreasing in t, where $t \geq s$. This result is intuitive because our model implies that the subject-specific event occurrence rate is positively correlated with the risk of failure event: subjects who survive longer tend to have lower event occurrence rates.

4 ESTIMATION PROCEDURE AND ASYMPTOTIC PROPERTIES

4.1 A BRIEF REVIEW

Let subscript i be the index for a subject, $i = 1, 2, \ldots, n$. For subject i, denote by X_i the time-independent covariate, Z_i the subject-specific latent variable, Y_i the observed terminating time for observing the event process $N_i(\cdot)$, D_i the failure time, and $\Delta_i = I(D_i \leq Y_i)$ the censoring indicator. We further denote by m_i the number of recurrent events occurred before

8
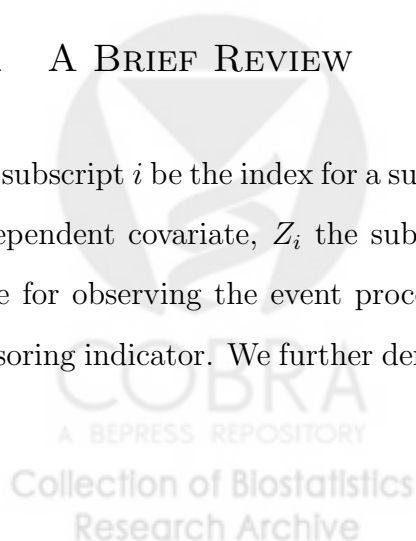
time Y_i and $t_{i1}, \ldots, t_{im_i}$ the observed event times for subject i. For ease of notation, we use m_i and t_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m_i$, to denote either random variables or realization values. Assume that $\{(X_i, Z_i, N_i(\cdot), D_i, C_i)\}$ are independent and identically distributed (iid), therefore the observed $\{(X_i, Z_i, m_i, (t_{i1}, \ldots, t_{im_i}), Y_i)\}$ are also iid.

Under assumption (M3), Y and $N(\cdot)$ are independent given the values of Z and X. The estimation procedure of Wang et al. (2001) can then be adopted to estimate Λ_0 and α. A key step of their estimation procedure is to observe that, conditional on (x_i, y_i, z_i, m_i), the observed event times, $\{t_{i1}, t_{i2}, \ldots, t_{im_i}\}$, are the order statistics of a set of iid random variables with the density function $\pi_i(t)$, where for $z_i > 0$

$$\pi_i(t) = \frac{z_i \lambda_0(t) \exp(\alpha x_i)}{z_i \Lambda_0(y_i) \exp(\alpha x_i)} = \frac{\lambda_0(t)}{\Lambda_0(y_i)}, \qquad 0 \le t \le y_i.$$

Note that $\pi_i(t)$ depends on neither z_i nor x_i, and it is a truncated density function of $\lambda_0(t)$ with observations truncated from the right side of y_i. As a result, the conditional likelihood function L_c given (x_i, y_i, z_i, m_i), where

$$L_c \propto \prod_{i=1}^{n} \prod_{j=1}^{m_i} \frac{\lambda_0(t_{ij})}{\Lambda_0(y_i)},$$

does not require information on x_i and the unobserved z_i. Although the data are correlated, computationally the conditional likelihood has the form of the nonparametric likelihood for independently right-truncated data. The nonparametric MLE of Λ_0, $\hat{\Lambda}_0$, based on randomly truncated data is known to have a product-limit representation,

$$\hat{\Lambda}_0(t) = \prod_{s_{(l)} > t} \left(1 - \frac{d_{(l)}}{R_{(l)}}\right),$$

where $\{s_{(l)}\}$ are the ordered and distinct values of the event times $\{t_{ij}\}$, $d_{(l)}$ is the number of events occurring at $s_{(l)}$, and $R_{(l)}$ is the total number of events with event time and observation terminating time satisfying $\{t_{ij} \le s_{(l)} \le y_i\}$.

It follows $E[m_i \mid X_i, Y_i, Z_i] = Z_i \exp(X_i \alpha) \Lambda_0(Y_i)$ that

$$E[m_i \Lambda_0^{-1}(Y_i) \mid X_i, Y_i] = E[E[m_i \mid X_i, Y_i, Z_i] \Lambda_0^{-1}(Y_i) \mid X_i, Y_i] = \mu_Z \exp(X_i \alpha),$$

9

Thus a class of estimating equations for α is defined as

$$n^{-1} \sum_{i=1}^{n} w_i \bar{X}_i^T (m_i \Lambda_0(Y_i)^{-1} - \exp(\bar{X}_i \gamma)) = 0, \tag{1}$$

where $\bar{X}_i = (1, X_i)$, $\gamma^T = (\ln(\mu_Z), \alpha^T)$, and w_i is a weight function depending on (X_i, γ, Λ_0). An estimate, $\hat{\alpha}$, of α can be obtained by solving the estimating equation with $\Lambda_0(Y_i)$ replaced by $\hat{\Lambda}_0(Y_i)$.

It is clearly seen that the estimation focus of Wang et al.(2001) was placed on the recurrent event process where the occurrence of the failure event is treated as a nuisance. In section 4.2 we shall consider inferential results for the failure event as well as the joint model.

4.2   A Borrow-Strength Method

Let $\hat{\mathcal{E}}$ and $\mathcal{E}$ represent the sample empirical means and the limit of average expectation, respectively. More specifically, for any function a of (X, Y, Z, Δ), let $\hat{\mathcal{E}}\{a(X, Y, Z, \Delta)\} = n^{-1} \sum_{i=1}^{n} a(X_i, Y_i, Z_i, \Delta_i)$ and $\mathcal{E}\{a(X, Y, Z, \Delta)\} = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} E[a(X_i, Y_i, Z_i, \Delta_i)]$, assuming existence of the limit.

Conditional on $\{(X_i, Y_i, Z_i), i = 1, \ldots, n\}$, under (M2) the score function derived from the partial likelihood can be expressed as

$$U(\beta) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \left\{ X_i - \frac{\sum_{j=1}^{n} X_j Z_j \exp(X_j \beta) \mathrm{I}(Y_j \geq Y_i)}{\sum_{j=1}^{n} Z_j \exp(X_j \beta) \mathrm{I}(Y_j \geq Y_i)} \right\} \mathrm{I}(Y_i \leq T_0)$$

$$= \hat{\mathcal{E}}\{X \Delta \mathrm{I}(Y \leq T_0)\} - \int_0^{T_0} \frac{\hat{\mathcal{E}}\{XZ \exp(X\beta) I(Y \geq s)\}}{\hat{\mathcal{E}}\{Z \exp(X\beta) I(Y \geq s)\}} d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\}. \tag{2}$$

U defines a functional of four empirical processes for each fixed β. It is known that, under mild regularity conditions, $U(\beta)$ converges almost surely to $\mathcal{U}(\beta)$ for each fixed β, where

$$\mathcal{U}(\beta) = \mathcal{E}\{X \Delta \mathrm{I}(Y \leq T_0)\} - \int_0^{T_0} \frac{\mathcal{E}\{XZ \exp(X\beta) I(Y \geq s)\}}{\mathcal{E}\{Z \exp(X\beta) I(Y \geq s)\}} d\mathcal{E}\{\Delta I(Y \leq s)\}.$$

10

Under (M3) and minor regularity conditions, it can be proved that the two equalities

$$d\mathcal{E}\{\Delta I(Y \leq s)\} = \mathcal{E}\{Z \exp(\beta X)I(Y \geq s)\}h_0(s)ds$$

and

$$d\mathcal{E}\{X\Delta I(Y \leq s)\} = \mathcal{E}\{XZ \exp(\beta X)I(Y \geq s)\}h_0(s)ds$$

hold when β satisfies (M2). It follows that $\mathcal{U}(\beta) = 0$ if β is the true regression parameter. By applying the Cauchy-Schwartz inequality to the derivative of $\mathcal{U}$, it can be further shown that the true regression parameter is the unique root (zero-crossing) of $\mathcal{U}$.

In reality, we are not able to observe the value of Z, and therefore cannot have the direct use of the score function U. Conditioning on (X_i, Y_i, Z_i), the expected value of m_i is $Z_i \exp(X_i\alpha)\Lambda_0(Y_i)$. It is natural to estimate Z_i by

$$\hat{Z}_i = \frac{m_i}{\hat{\Lambda}_0(Y_i)e^{X_i\hat{\alpha}}} \ ,$$

where $\hat{\Lambda}_0(\cdot)$ and $\hat{\alpha}$ are obtained from the estimation procedure discussed in the previous section. We propose a "borrow-strength estimation procedure" as follows: First, compute the individual frailty value $\hat{Z}_i$. Next, estimate the empirical processes in the score function (2) by plugging in $(\hat{Z}_1, \ldots, \hat{Z}_n)$, and, in the final step, use this working score function to estimate β.

Note that the estimate of $\hat{\Lambda}_0(t)$, and hence $\hat{Z}_i$, is obtained from the entire collection of recurrent event data, and $\hat{Z}_i$ captures the subject-specific characteristics under Model (M1~3). The proposed estimator $\hat{Z}_i$ possesses desirable moment properties: the two processes $\hat{\mathcal{E}}\{\hat{Z} \exp(X\beta)I(Y \geq s)\}$ and $\hat{\mathcal{E}}\{X\hat{Z} \exp(X\beta)I(Y \geq s)\}$ will be shown in the next section to converge almost surely to the limits $\mathcal{E}\{Z \exp(X\beta)I(Y \geq s)\}$ and $\mathcal{E}\{XZ \exp(X\beta)I(Y \geq s)\}$, respectively, for each fixed β. Therefore this strength-borrowing method allows the working score function to attain the same limit $\mathcal{U}$ as if the latent variable were observed. The zero-crossing of the working score function serves as an estimator of the zero-crossing of $\mathcal{U}$,

11

that is, β. To be specific, the working score function $\hat{U}$ of U is given by

$$\hat{U}(\beta) = \hat{\mathcal{E}}\{X\Delta\mathrm{I}(Y \leq T_0)\} - \int_0^{T_0} \frac{\hat{\mathcal{E}}\{X\hat{Z}\exp(X\beta)I(Y \geq s)\}}{\hat{\mathcal{E}}\{\hat{Z}\exp(X\beta)I(Y \geq s)\}} d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\}, \qquad (3)$$

with the usual convention that $0/0 = 0$. It will be shown in Section 5 that $\hat{U}$ converges to $\mathcal{U}$ almost surely in a neighborhood of β. We then estimate β by $\hat{\beta}$, where $\hat{U}(\hat{\beta}) = 0$.

If Z were observed, the Breslow estimator $\tilde{H}_0(t)$ of the baseline cumulative hazard function, H_0, would be

$$\tilde{H}_0(t) = \int_0^t \frac{d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\}}{\hat{\mathcal{E}}\{Z\exp(X\hat{\beta})I(Y \geq s)\}},$$

which is a functional of two empirical processes. Under the conditional independence assumption of C and D, given (X, Z), we can show that the baseline cumulative hazard function, $H_0(t)$, is the limit of $\tilde{H}_0(t)$.

As with the estimation procedure for the regression parameters, we propose an estimator of $H_0(t)$ as

$$\hat{H}_0(t) = \int_0^t \frac{d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\}}{\hat{\mathcal{E}}\{\hat{Z}\exp(X\hat{\beta})I(Y \geq s)\}}. \qquad (4)$$

The limit of the estimator $\hat{H}_0(t)$ can be shown to be the functional of the limits of the two processes in (4), i.e., $\hat{H}_0(t) \to H_0(t)$ almost surely. The asymptotic normality of the proposed estimator $\hat{H}_0(t)$ will be studied in the next section.

5 LARGE SAMPLE PROPERTIES

In order to study the large sample properties of the proposed estimators, we impose the following regularity conditions:

(A1) $\Pr(Y \geq T_0, Z > 0) > 0$,

12
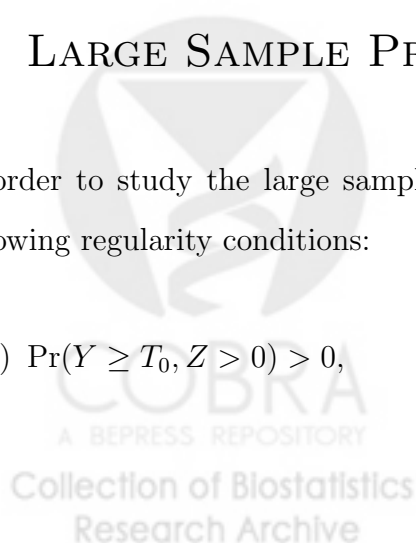
(A2) X is uniformly bounded,

(A3) $EZ^2 < \infty$, and

(A4) $G(u) = E[ZI(Y \geq u)]$ is a continuous function for $u \in [0, T_0]$.

Under these regularity conditions, the large sample properties of $\hat{\Lambda}_0$ and $\hat{\alpha}$ were established in Wang et al. (2001). We denote the following asymptotic representations:
$\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t)) = n^{-1/2} \sum_{i=1}^{n} \Lambda_0(t)b_i(t) + o_p(1)$, **for** $\inf\{y : \Lambda_0(y) > 0\} < t < T_0$,
and, provided $E[\partial e_i / \partial \gamma]$ is nonsingular, $\sqrt{n}(\hat{\alpha} - \alpha) = n^{-1/2} \sum_{i=1}^{n} f_i(\alpha) + o_p(1)$, where
$\gamma^T = (\ln(\mu_Z), \alpha^T)$ and $b_i(t)$, f_i, and e_i are defined in Appendix B. Note that in the WQC
model, the baseline cumulative intensity function was not assumed to satisfy $\Lambda_0(T_0) = 1$ as
we assumed in (M1). The forementioned asymptotic representations have been modified in
order to accommodate the current model assumptions.

The weak convergence of $\sqrt{n}(\hat{\mathcal{E}}\{X\Delta I(Y \leq T_0)\} - \mathcal{E}\{X\Delta I(Y \leq T_0)\})$ and $\sqrt{n}(\hat{\mathcal{E}}\{\Delta I(Y \leq t)\} - \mathcal{E}\{\Delta I(Y \leq t)\})$ follow from the classical central limit theorem and Example 2.11.16
of van der Vaart and Wellner (1996). The two empirical processes converge weakly to a
zero-mean normal distribution, $\mathcal{W}_1$, and a zero-mean Gaussian process, $\mathcal{W}_2$, respectively.

Furthermore, by denoting V the joint probability density function of (X, Y, m) and
arguing as in the proof of Theorem 1 in Wang et al. (2001), we are able to show that
$\sqrt{n}\big(\hat{\mathcal{E}}\{\hat{Z}\exp(Xb)I(Y \geq t)\} - \mathcal{E}\{Z\exp(Xb)I(Y \geq t)\}\big) = n^{-1/2} \sum_{i=1}^{n} \psi_{3i}(t; b) + o_p(1)$, where

$$\psi_{3i}(t; b) = \int m\Lambda_0^{-1}(y)e^{x(b-\alpha)}I(y \geq t)(xf_i(\alpha) + b_i(y))dV(x, y, m)$$

$$+ m_i\Lambda_0^{-1}(Y_i)e^{X_i(b-\alpha)}I(Y_i \geq t) - \mathcal{E}\{Ze^{Xb}I(Y \geq t)\},$$

with the usual convention $0/0 = 0$. Note that ψ_{3i}'s are uncorrelated random variables since
$\psi_{3i}(t; b)$ depends only on observed data from the ith individual. It follows the law of large
numbers that $\hat{\mathcal{E}}\{\hat{Z}\exp(Xb)I(Y \geq t)\} - \mathcal{E}\{Z\exp(Xb)I(Y \geq t)\} \to 0$ almost surely, for each

13

fixed b. Furthermore, by the central limit theorem, the process converges in finite dimension to a zero-mean Gaussian process $\mathcal{W}_3$ on the time interval $[0, T_0]$. The explanatory variable X is assumed to be bounded, without loss of generality we assume that $X \geq 0$. Because items in $\psi_{3i}(t; b)$ are monotone processes for each b, the process $\psi_{3i}(t; b)$ is tight and converges weakly to $\mathcal{W}_3$ (see Example 2.11.16 of van der Vaart and Wellner 1996). Similar arguments hold for $\hat{\mathcal{E}}\{X\hat{Z}\exp(Xb)I(Y \geq t)\} - \mathcal{E}\{XZ\exp(Xb)I(Y \geq t)\} \to 0$ almost surely, and the process $\sqrt{n}(\hat{\mathcal{E}}\{X\hat{Z}\exp(Xb)I(Y \geq t)\} - \mathcal{E}\{X\hat{Z}\exp(Xb)I(Y \geq t)\})$ has the asymptotically iid representation $n^{-1/2}\sum_{i=1}^n \psi_{4i}(t; b) + o_p(1)$, where ψ_{4i} is defined by

$$\psi_{4i}(t; b) = \int mx\Lambda_0^{-1}(y)e^{x(b-\alpha)}I(y \geq t)(xf_i(\alpha) + b_i(y))dV(x, y, m)$$

$$+m_ix_i\Lambda_0^{-1}(Y_i)e^{X_i(b-\alpha)}I(Y_i \geq t) - \mathcal{E}\{XZe^{Xb}I(Y \geq t)\}.$$

Moreover, the process converges weakly to a zero-mean Gaussian process, denoted by $\mathcal{W}_4$.

We establish the consistency of $\hat{\beta}$ as follows. Define the two functions

$$A_n(b) = \hat{\mathcal{E}}\{(X\Delta I(Y \leq T_0)\}(b - \beta) - \int_0^{T_0} \ln\Big(\frac{\hat{\mathcal{E}}\{\hat{Z}\exp(Xb)I(Y \geq s)\}}{\hat{\mathcal{E}}\{\hat{Z}\exp(X\beta)I(Y \geq s)\}}\Big)d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\},$$

and

$$A(b) = \mathcal{E}\{(X\Delta I(Y \leq T_0)\}(b - \beta) - \int_0^{T_0} \ln\Big(\frac{\mathcal{E}\{Z\exp(Xb)I(Y \geq s)\}}{\mathcal{E}\{Z\exp(X\beta)I(Y \geq s)\}}\Big)d\mathcal{E}\{\Delta I(Y \leq s)\}.$$

One can easily verify that $\hat{U}(b)$ and $\mathcal{U}(b)$ are derivatives of $A_n(b)$ and $A(b)$, respectively, and β is the unique maximum of A. Furthermore, $\hat{\beta}$ can be shown to be the unique maximum of A_n.

From the foregoing discussions, the four processes in $\hat{U}$ has the $\sqrt{n}$-convergence rate, hence the four processes converge almost surely to their limits. Applying Lemma 3 of Gill (1989) and the chain rule, one can show that the functional defined by $\hat{U}$ is continuous with respect to the supremum norm under regularity conditions (A1~4). Then, for some

14

compact neighborhood $\mathcal{B}$ of β, as $n \to \infty$, $\sup_{b \in \mathcal{B}} | \hat{U}(b) - \mathcal{U}(b) | \to 0$ almost surely. Apply Taylor expansion and by the fact that $A_n(\beta) = A(\beta) = 0$, one has $A_n(b) - A(b) = \{\hat{U}(\beta^*) - \mathcal{U}(\beta^*)\}(b - \beta)$, where β^* lies between b and β. Now it is clear that, for $n \to \infty$, $\sup_{b \in \mathcal{B}} | A_n(b) - A(b) | \to 0$ almost surely.

Define $\hat{\Gamma}(b) = d\hat{U}(b)/db = d^2 A_n(b)/db^2$ and $\hat{\Gamma}(b) = d\mathcal{U}(b)/db = d^2 A(b)/db^2$, that is,

$$\hat{\Gamma}(b) = \int_0^{T_0} -\frac{\hat{\mathcal{E}}\{X^2 \hat{Z} e^{Xb} I(Y \geq s)\}}{\hat{\mathcal{E}}\{\hat{Z} e^{Xb} I(Y \geq s)\}} + \frac{\hat{\mathcal{E}}\{X \hat{Z} e^{Xb} I(Y \geq s)\}^2}{\hat{\mathcal{E}}\{\hat{Z} e^{Xb} I(Y \geq s)\}^2} d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\},$$

and

$$\Gamma(b) = \int_0^{T_0} -\frac{\mathcal{E}\{X^2 Z e^{Xb} I(Y \geq s)\}}{\mathcal{E}\{Z e^{Xb} I(Y \geq s)\}} + \frac{\mathcal{E}\{X Z e^{Xb} I(Y \geq s)\}^2}{\mathcal{E}\{Z e^{Xb} I(Y \geq s)\}^2} d\mathcal{E}\{\Delta I(Y \leq s)\}$$

One can show that $\hat{\Gamma}(b)$ and $\Gamma(b)$ are both negative definite, and it follows that A_n and A are concave. By Lenglart's theorem (Appendix II in Andersen and Gill 1982), the unique maximum of A_n, $\hat{\beta}$, converges in probability to the unique maximum of A, i.e., β. Hence we establish the consistency of $\hat{\beta}$.

Note that $\{\hat{Z}_1, \ldots, \hat{Z}_n\}$ are correlated because these values are estimated from the entire collection of recurrent event data; therefore, martingale theory does not apply to the working score function, $\hat{U}$. In this paper the large sample properties of $\hat{\beta}$ and $\hat{H}_0(t)$ are studied by empirical process theories and the functional delta method. For convenience we denote $a^2 = aa^T$ for any vector a. We present asymptotic theories in Lemmas 1 through 3, with proofs given in Appendix B, and summarize these results in Theorem 1.

Lemma 1. Under regularity conditions (A1~4) and the assumption that $\Psi = E[\partial e_1/\partial \gamma]$ is nonsingular, $n^{1/2}\hat{U}(\beta)$ is the sum of asymptotically uncorrelated random variables: $n^{1/2}\hat{U}(\beta) = n^{-1/2} \sum_{i=1}^n \psi_i(\beta) + o_p(1)$, where $\psi_i(\beta)$ is defined in Appendix B. Moreover, $\sqrt{n}\hat{U}(\beta)$ converges weakly to a normal distribution with zero mean and variance-covariance matrix $\Sigma(\beta) = E[\psi_i(\beta)^2]$.

15

Note that the variance-covariance matrix Σ can be consistently estimated by $\hat{\Sigma}(\hat{\beta})$, where $\hat{\Sigma}(\hat{\beta})$ is defined in Appendix B. To study the large sample property of $\hat{\beta}$, we further define $\Gamma(\beta) = \partial\mathcal{U}(\beta)/\partial\beta$, and $\hat{\Gamma}(\beta) = \partial\hat{U}(\beta)/\partial\beta$.

Lemma 2. Assume that Ψ and $\Gamma = \Gamma(\beta)$ are both nonsingular. Then, under regularity conditions (A1~4), $\sqrt{n}(\hat{\beta} - \beta) = n^{-1/2}\sum_{i=1}^{n}\Gamma^{-1}\psi_i(\beta) + o_p(1)$, where $\psi(\beta)$ is defined in Appendix B. Thus $\sqrt{n}(\hat{\beta} - \beta)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Gamma^{-1}\Sigma(\Gamma^{-1})^T$, which can be consistently estimated by $\hat{\Gamma}(\hat{\beta})^{-1}\hat{\Sigma}(\hat{\beta})\{\hat{\Gamma}(\hat{\beta})^{-1}\}^T$.

Lemma 3. Under regularity conditions (A1~4) and by assuming that Ψ and Γ are nonsingular, the cumulative hazard function, $H_0(t)$, can be expressed as the sum of asymptotically uncorrelated random variables: $n^{1/2}\{\hat{H}_0(t) - H_0(t)\} = n^{-1/2}\sum_{i=1}^{n}\phi_i(t) + o_p(1)$, where $t \in [0, T_0]$ and $\phi_i(t)$ is defined in Appendix B. Then $n^{1/2}\{\hat{H}_0(t) - H_0(t)\}$ converges weakly on $[0, T_0]$ to a zero mean Gaussian process with variance-covariance function $E[\phi_1(t_1)\phi_1(t_2)]$.

Along with the results stated in Section 4.1 and following directly from Lemmas 2 and 3, we state the main asymptotic theorem below:

Theorem 1. Assume that Γ and Ψ are nonsingular. Under regularity conditions (A1~4), for each fixed s, $\inf\{\Lambda_0(y) > 0\} < s < T_0$, and fixed t, $t \in [0, T_0]$, the random vector $\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta, \hat{\Lambda}_0(s) - \Lambda_0(s), \hat{H}_0(t) - H_0(t))$ converges weakly to a multivariate normal distribution with mean 0 and variance covariance matrix $E[\eta_1^2]$, where η_i's are uncorrelated random vectors defined by $\eta_i = (f_i(\alpha),\ \Gamma\psi_i(\beta),\ \Lambda_0(s)b_i(s),\ \phi_i(t))$.

16

6 SIMULATIONS AND DATA ANALYSIS

6.1 MONTE CARLO SIMULATIONS

Simulation studies have been conducted to assess the performance of the proposed estimators. For all simulation studies, 1,000 simulated data sets are generated, each with $n = 200$ and $n = 500$ independent subjects. The explanatory variable X was generated from a Bernoulli distribution with $P(X = 0) = P(X = 1) = 0.5$, and the subject specific latent variable Z was generated from a discrete (poisson with mean 10) and a continuous (gamma with mean 10 and variance 50) distribution. Given $X = x$ and $Z = z$, the subject's underlying recurrent event process $\{N(t),\ t \in [0, 10]\}$ is a non-stationary Poisson process with the corresponding intensity function $z\lambda_0(t)\exp(x\alpha)$ and the subject's failure time D has a hazard function $zh_0(t)\exp(x\beta)$. To examine the performance of proposed estimators under different choices of (α, β) and $(\lambda_0(\cdot), h_0(\cdot))$, we also consider combinations corresponding to $(\alpha, \beta) = (0, 0)$ and $(-1, -1.5)$ and the following two sets of functions for $\lambda_0(t)$ and $h_0(t)$. Scenario I: $\lambda_0(t) = 1/10$, and $h_0(t) = t/400$; Scenario II: $h_0(t) = (t + 1)/10$, $h_0(t) = \sqrt{t}/200$. Finally, the censoring time C is either a exponential variable with mean 10 when $x = 1$, or a exponential variable with mean $300/z^2$ when $x = 0$. Given (x, z), the triplets $(N(\cdot), D, C)$ are mutually independent.

Suppose the censoring time C is the potential dropout time. The justification of such a design for the censoring variable is the following: Suppose the frailty is an unobserved health indicator, in the control group $(X = 0)$, sick patients with high occurrence rate of recurrent events drop out early due to large values of frailty; in the treatment group $(X = 1)$, on the other hand, because the treatment has effectively reduced the event occurrence rates, the dropout is non-informative for both the recurrent event process and the failure time.

As summarized in Table 1, the average death rate ranges from 13% to 28%, the aver-

17

age length of follow-up time ranges from 3.9 to 4.91, and the average number of observed recurrent events ranges from 1.57 to 3.65 in the conducted simulation studies. It is noted that the average follow-up time is approximately the same under different choices of (λ_0, h_0), but the average number of observed events is smaller under Scenario II. The result of simulation studies is summarized in Table 2. For each simulation study, the empirical bias, standard error, and correlation coefficient of proposed estimators were calculated based on 1,000 samples. Figures 1 and 2 show the estimates and the pointwise 95% confidence intervals of the baseline cumulative intensity function and baseline cumulative hazard function. As shown in Table 2 and Figures 1 and 2, the proposed estimator performs reasonably well; that is, the empirical bias in the estimates of regression parameters are small and the averages of $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ are almost indistinguishable from the true curves. Note that the parameter estimates under $Z \sim poisson(10)$ have smaller standard errors than those under $Z \sim gamma(2, 5)$, and the empirical correlation coefficients between $\hat{\alpha}$ and $\hat{\beta}$ are smaller under the assumed poisson distribution; this is because the $poisson(10)$ distribution has smaller variability than the $gamma(2, 5)$ distribution, i.e., the defined population is more homogeneous under $Z \sim poisson(10)$.

With data generated by Model (M1~3), it is interesting to see results from the use of a popular but incorrect model, i.e., the proportional hazards model, $h(t) = h_0^*(t) \exp(x\beta^*)$, for the failure time data. By using the partial likelihood method (Cox, 1972), Table 2 also reports the average and empirical standard error of the 1000 estimates of β^*. It is observed that using the Cox proportional hazards model, which incorrectly assumes the independent censoring assumption, results in biased estimation of the treatment effect. This phenomenon can be explained as follows: In the simulated control group $(X = 0)$ sicker patients with higher hazards tend to drop out at earlier times, therefore, risk sets are likely to consist of healthier patients at later time points. As a result, the estimates given by the Cox proportional hazards model based on comparisons of subjects within risk sets underestimates

18

the treatment effect when treatment reduces the risk of death and conclude that treatment is associate with increased risk of death when the treatment does not affect the mortality rate.

6.2 DATA ANALYSIS

A Denmark registry data set recorded the initial and recurrences of hospitalizations and associated patient information from 8,811 patients whose first schizophrenia-related hospitalization occurred between April 1, 1970 and March 25, 1988 (Eaton et al. 1992a, b). The catchment area for the register is the entire nation of Denmark. The data provide a large collection of repeated psychiatric measurements as well as recorded hospitalization episodes. All death records in Denmark are linked into to the register.

Table 3 summarizes numbers of hospital admissions and deaths for subgroups by gender and age of onset. Comparing crude proportions seems to suggest that patients whose first hospitalization occurred after age 20 tend to have fewer hospitalizations but are more likely to die before the end of study. The hospitalizations and survival experiences do not look very different in males and females based on these summary statistics.

We apply the proposed joint model to the Denmark schizophrenia cohort data and investigate the effects of gender and age of onset on the rate of hospitalization and the risk of death. The gender indicator is set to be 1 for male and 0 for female, and the indicator of age onset is set to be 1 if less than 20 years of age and 0 if greater than or equal to 20.

To estimate the standard errors of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ at selected time points, a nonparametric bootstrap method for clustered data was adopted by repeatedly sampling 8811 subjects with replacement, using subject as the sampling unit, from the schizophrenia cohort data. The results of data analysis are summarized in Table 4. Estimates of $\Lambda_0(t)$ and $H_0(t)$, and their pointwise 95% bootstrap confidence intervals are given in Figure 3. Table

19

4 shows that patients with early onset (≤ 20 y/o) are hospitalized more often (21% higher) and have decreased risk of death (57% lower). Moreover, being a male will decrease one's occurrence rate of hospitalization episodes and the risk of death by 16% ($\approx 1 - e^{-0.18}$) and 10% ($\approx 1 - e^{-0.11}$), respectively. The estimated covariate effects are statistically significant, except for the gender effect on the risk of death, which is marginally significant. It is interesting to see that the age of onset has opposite effects on hospital admissions rate and the hazard of death - this is not surprising, however, since young patients tend to have longer life expectancy. Also, the analysis confirms the theory in schizophrenia that patients with early onset age tend to be hospitalized more often than those with later onset age.

In the case of a degenerate frailty, the Cox proportional hazards model gives estimates of -0.74 (S.E.$= 0.12$) and -0.14 (S.E.$= 0.06$) for the effects of early onset and gender, respectively. The direction of covariate effects estimated in the Cox proportional hazards model are consistent with the estimates under the proposed model.

7 DISCUSSIONS

Frailty models are commonly adopted in modeling multivariate survival time data (Clayton, 1978; Oakes 1982) and in jointly modeling repeated measures and survival time data (Henderson, Diggle, and Dobson, 2000; Lin, Turnbull, McCulloch, and Slate, 2002). In this paper, we propose a semiparametric joint model for the recurrent event process and failure time data. A latent variable (frailty) is assumed to act as a multiplicative factor in both the intensity function and the hazard function, and hence induces the correlation between the event process and the failure time. Unlike the usual setting of frailty models where a parametric distribution is assumed for the frailty, a specific feature of our model is that the frailty distribution is treated as nuisances and no parametric assumptions were imposed. Additionally, via the use of frailty, the proposed model relaxes the independent censoring

20

condition for observing both the recurrent event process and the failure time data.

For a semiparametric model like (M1∼3), model checking is expected to be a difficult task in general. This paper does not intend to develop methods for formal model checking and we simply suggest possible approaches for the validation of model assumptions. Rigorous study of model checking methods will be conducted elsewhere. To test the assumption of a common baseline intensity function shared by all subjects, we utilize the fact that, under (M1) and conditioning on (m_i, x_i, y_i, z_i), t_{ij} are iid with the cdf $F(t)I(0 \leq t \leq y_i)/F(y_i)$. Define $V_{ij} = F(t_{ij})I(0 \leq t_{ij} \leq y_i)/F(y_i)$ then V_{ij} are order statistics of iid uniform$(0, 1)$ random variables. Let $\hat{V}_{ij} = \hat{F}(t_{ij})I(0 \leq t_{ij} \leq y_i)/\hat{F}(y_i)$, then a necessary condition to validate the assumption of sharing a common intensity function is to check if the empirical distribution of $\{\hat{V}_{ij} : j = 1, \ldots, m_i; i = 1 \ldots, n\}$ is approximately uniform$(0, 1)$ distribution. To check on the proportional rate and hazards model assumption imposed by (M1) and (M2), respectively, replace Z with $\hat{Z}$ to derive the Schoenfeld residuals (Schoenfeld 1982). If the assumption of proportional hazards holds, the derived residuals are expected to randomly fluctuate around zero.

In this paper, a borrow-strength procedure was proposed by first estimating the value of the latent variable from recurrent event data, and next using the estimated value in the failure time models. The central idea of estimation is to utilize moment properties of $\hat{Z}$ so that the partial score functions, with Z or $\hat{Z}$, attain the same convergence function. The proposed $\hat{Z}$ requires no parametric assumption on Z and is easy to compute. As opposed to this approach, an alternative choice is to estimate Z by the posterior mean of Z given the observed recurrent event data; however, as discussed in Implication 1, the posterior mean does not have an explicit form in our model setting, and is therefore not a useful choice in theory or application.

The proposed estimation procedure is not without constraints - it is applicable only to

21

time-independent covariates. In some applications, it would be desirable to develop estimation procedures that allow for both time-invariant and time-dependent covariates. Also, the implications and trajectories described in Section 3 help understand the general relationship between the recurrent event process and the failure time. However, the probability formulas established in Section 3 can not be made explicit unless the unknown parameters in the formulas are known or estimable, and, to accomplish such a task, it requires more parametric modeling and alternative estimation procedures. Such work will be considered elsewhere in the future. Finally, the proposed time-to-events models assume a common baseline intensity/rate function be shared by all subjects and the intensity/rate function does not change after the occurrence of an event. To characterize the possible change in the risk of event occurrence after each event time, techniques for time-between-events models by, say, Prentice, Williams, and Peterson (1981) and Chang and Wang (1999) can be adopted.

Appendix A

Implication 1: The probability density function of the event history given the value of the frailty and the termination time can be expressed as

$$f(\mathcal{H}(y) \mid z, y) = f(\mathcal{H}(y) \mid N(y), z, y) f(N(y) \mid z, y) = \left(\prod_{j=1}^{N(y)} \frac{\lambda_0(t_j)}{\Lambda_0(y)} \right) f(N(y) \mid z, y),$$

where $f(N(y) \mid z, y)$ is the probability density function of the number of observed recurrent events given the value of the frailty and the termination time. Consequently,

$$f(\mathcal{H}(y) \mid y) = \int f(\mathcal{H}(y) \mid z, y) f(z \mid y) dz = \left(\prod_{j=1}^{N(y)} \frac{\lambda_0(t_j)}{\Lambda_0(y)} \right) f(N(y) \mid y) .$$

Thus we can write the posterior mean of Z, given the observed recurrent event data, as

$$E[Z \mid \mathcal{H}(y), y] = \int z f(\mathcal{H}(y) \mid z, y) \frac{f(z \mid y)}{f(\mathcal{H}(y) \mid y)} dz = \int z f(N(y) \mid z, y) \times \frac{f(z \mid y)}{f(N(y) \mid y)} dz$$

$$= \int z \times \frac{\exp(-z\Lambda_0(y))(z\Lambda_0(y))^{N(y)}}{N(y)!} \times \frac{f(z \mid y)}{f(N(y) \mid y)} dz$$

$$= \quad \frac{N(y)+1}{\Lambda_0(y)} \int f(N(y)+1 \mid z, y) \frac{f(z \mid y)}{f(N(y) \mid y)} dz$$

$$= \quad \frac{N(y)+1}{\Lambda_0(y)} \times \frac{f(N(y)+1 \mid y)}{f(N(y) \mid y)} .$$

Implication 2: For $0 \leq t \leq t+s \leq T_0$, the survival function of the residual life time after time t, given the event history before and up to time t, can be expressed as $P(D \geq t+s \mid \mathcal{H}(t), D \geq t) = P(D \geq t+s, \mathcal{H}(t))/P(D \geq t, \mathcal{H}(t))$, where

$$P(D \geq t+s, \mathcal{H}(t)) \quad = \quad \int P(D \geq t+s \mid z) f(\mathcal{H}(t) \mid z) f_Z(z) dz$$

$$= \quad \int P(D \geq t+s \mid z) f(\mathcal{H}(t) \mid N(t), z) f(N(t) \mid z) f_Z(z) dz$$

$$= \quad \int e^{-z H_0(t+s)} \times \left(\prod_{j=1}^{N(t)} \frac{\lambda_0(t_j)}{\Lambda_0(t)} \right) \times \frac{e^{-z\Lambda_0(t)}(z\Lambda_0(t))^{N(t)}}{N(t)!} f_Z(z) dz$$

$$= \quad \frac{1}{N(t)!} \prod_{j=1}^{N(t)} \lambda_0(t_j) \times E[e^{-Z\{H_0(t+s)+\Lambda_0(t)\}} Z^{N(t)}],$$

and, similarly, $P(D \geq t, \mathcal{H}(t)) = (N(t)!)^{-1} \prod_{j=1}^{N(t)} \lambda_0(t_j) \times E[e^{-Z\{H_0(t)+\Lambda_0(t)Z^{N(t)}\}}]$. We then simplify the formula

$$P(D \geq t+s \mid \mathcal{H}(t), D \geq t) \quad = \quad \frac{E[e^{-Z\{H_0(t+s)+\Lambda_0(t)\}} Z^{N(t)}]}{E[e^{-Z\{H_0(t)+\Lambda_0(t)\}} Z^{N(t)}]} .$$

Implication 4: Following (M3), the mean function of the recurrent event conditional on the failure time can be expressed as

$$E[N(s) \mid D \geq t] \quad = \quad \int \frac{E[N(s) \mid z] \, P(D \geq t \mid z)}{P(D \geq t)} f_Z(z) dz$$

$$= \quad \int \frac{z\Lambda_0(s) \times e^{-z H_0(t)}}{E[e^{-Z H_0(t)}]} f_z(z) dz = \frac{E[Z e^{-Z H_0(t)}]}{E[e^{-Z H_0(t)}]} \Lambda_0(s). \tag{5}$$

23
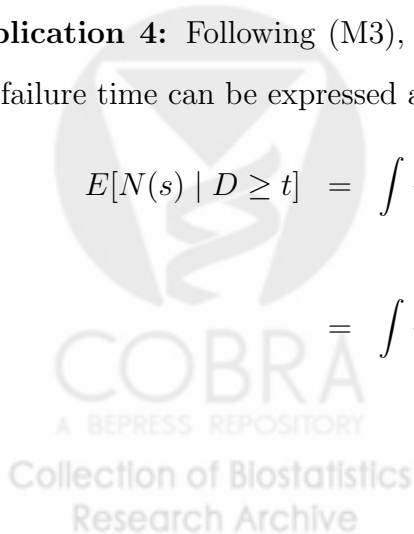
The partial derivative of the right-hand-side term in (5) with respect to t can be derived as

$$\frac{E[Ze^{-ZH_0(t)}]^2 - E[Z^2 e^{-ZH_0(t)}]E[e^{-ZH_0(t)}]}{E[e^{-ZH_0(t)}]^2} h_0(t)\Lambda_0(s).$$

The partial derivative can be shown to be nonpositive by applying the Cauchy-Schwartz inequality, and, as a result, the mean function in (5) is decreasing in t, $t \geq s$.

Appendix B

Denote t_{ij} the jth event time of the ith subject, and define the functions $G(t) = E[Z_1 I(Y_1 \geq t)]$, $R(t) = G(t)\Lambda_0(t)$, $Q(t) = \int_0^t G(u)d\Lambda_0(u)$, and, for $i = 1, \ldots, n$,

$$b_i(t) = \sum_{i=1}^{m_i} \Big\{ \int_t^{T_0} \frac{I(t_{ij} \leq u \leq Y_i)dQ(u)}{R^2(u)} - \frac{I(t < t_{ij} \leq T_0)}{R(t_{ij})} \Big\}.$$

Under regularity conditions (A1~4), it has been shown in Wang et al. (2001) that $\hat{\Lambda}_0(t) - \Lambda_0(t) = n^{-1}\sum_{i=1}^n \Lambda_0(t)b_i(t) + o_p(n^{-1/2})$, for $\inf\{\Lambda_0(y) > 0\} < t < T_0$, and $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ converges weakly to a normal distribution with zero mean and variance $\Lambda_0(t)^2 E[b_1^2(t)]$.

Define V^* to be the joint probability measure of $(w, \bar{X}, m, Y)$ and

$$e_i = -\int \frac{w\bar{x}^t m b_i(y)}{\Lambda_0(y)} dV^*(w, \bar{x}, m, y) + w_i \bar{x}_i^t \Big\{ m_i \Lambda_0(y_i)^{-1} - \exp(\bar{x}_i\gamma) \Big\}.$$

Then the left-hand-side of the estimating function (1) can be expressed as $n^{-1}\sum_{i=1}^n e_i + o_p(n^{-1/2})$. Assuming $E[\partial e_1/\partial \gamma]$ is nonsingular, one has $\sqrt{n}(\hat{\alpha} - \alpha) = n^{-1/2}\sum_{i=1}^n f_i(\alpha) + o_p(1)$, where $f_i(\alpha)$ is the vector function $E[-\partial e_1/\partial \gamma]^{-1}e_i$ without the first entry, and $\sqrt{n}(\hat{\alpha} - \alpha)$ converges to a multivariate normal distribution with zero mean and variance $E[f_1^2]$.

Proof of Lemma 1.

Straightforward algebra yields

$$n^{1/2}\hat{U}(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^{T_0} X_i - \frac{\hat{\mathcal{E}}\{X\hat{Z}e^{X\beta}I(Y \geq s)\}}{\hat{\mathcal{E}}\{\hat{Z}e^{X\beta}I(Y \geq s)\}} \Big\{ d\Delta_i I(Y_i \leq s) - \hat{Z}_i e^{X_i\beta}I(Y_i \geq s)h_0(s)ds \Big\}.$$

24

Since the mapping of $\hat{U}$ from the four empirical processes, under the regularity conditions, is compactly differentiable with respect to the supremum norm and the four empirical processes converge weakly to their limits, we apply the functional delta method to $\hat{U}$ and establish its asymptotic representation $n^{1/2}\hat{U}(\beta) = n^{-1/2}\sum_{i=1}^{n}\psi_i(\beta) + o_p(1)$, where

$$\psi_i(\beta) = X_i\Delta_i I(Y_i \leq T_0) - \mathcal{E}\{X\Delta I(Y \leq T_0)\}$$

$$+ \int_0^{T_0} \frac{\psi_{3i}(s;\beta)\mathcal{E}\{XZe^{X\beta}I(Y \geq s)\}}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}^2} d\mathcal{E}\{\Delta I(Y \leq s)\}$$

$$- \int_0^{T_0} \frac{\psi_{4i}(s;\beta)}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}} d\mathcal{E}\{\Delta I(Y \leq s)\}$$

$$- \int_0^{T_0} \frac{\mathcal{E}\{XZe^{X\beta}I(Y \geq s)\}}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}} d\Big(\Delta_i I(Y_i \leq s) - \mathcal{E}\{\Delta I(Y \leq s)\}\Big).$$

Note that ψ_i's are uncorrelated random variables because ψ_i depends only on the observed data of the ith individual. Following the classical central limit theorem, $n^{1/2}\hat{U}(\beta)$ is asymptotically normal with zero mean and variance-covariance matrix $\Sigma(\beta) = E[\psi_i(\beta)^2]$. Define $\hat{\psi}_i(\beta)$ by substituting empirical processes for their limits in ψ_i, and define $\hat{\Sigma}(\beta) = n^{-1}\sum_{i=1}^{n}\{\hat{\psi}_i(\beta) - \psi^*(\beta)\}\{\hat{\psi}_i(\beta) - \psi^*(\beta)\}^T$, where $\psi^*(\beta)$ is the average over $\hat{\psi}_1(\beta), \ldots, \hat{\psi}_n(\beta)$. It can be shown that the second moment of $\hat{\psi}_i(\beta)$ exists, and it follows from the strong law of large numbers that $\hat{\Sigma}(\beta)$ converges to its limit, $\Sigma(\beta)$, uniformly. Arguing as in the proof for the consistency of $\hat{\beta}$, we can show that the functional defined by $\hat{\Sigma}$ satisfies $\sup_{b \in \mathcal{B}} |\hat{\Sigma}(b) - \Sigma(b)| \to 0$ almost surely. By the consistency of $\hat{\beta}$, as well as the continuity of $\Sigma(b)$ at β, we are able to show that $\hat{\Sigma}(\hat{\beta})$ is a consistent estimator of $\Sigma(\beta)$. Moreover, in terms of the notations used before, we can rewrite the limit of $\sqrt{n}\hat{U}(\beta)$ as

$$\mathcal{W}_1 - \int_0^{T_0} \mathcal{W}_4(s)h_0(s)ds - \int_0^{T_0} \frac{\mathcal{E}\{XZe^{X\beta}I(Y \geq s)\}}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}}\{d\mathcal{W}_2(s) - \mathcal{W}_3(s)h_0(s)ds\}.$$

Proof of Lemma 2.

25

Define $\hat{\Gamma}(b) = d\hat{U}(b)/db$, that is,

$$\hat{\Gamma}(b) = \int_0^{T_0} -\frac{\hat{\mathcal{E}}\{X^2\hat{Z}e^{Xb}I(Y \geq s)\}}{\hat{\mathcal{E}}\{\hat{Z}e^{Xb}I(Y \geq s)\}} + \frac{\hat{\mathcal{E}}\{X\hat{Z}e^{Xb}I(Y \geq s)\}^2}{\hat{\mathcal{E}}\{\hat{Z}e^{Xb}I(Y \geq s)\}^2} d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\}.$$

It can be shown that $\hat{\Gamma}(b)$ defines a functional of four empirical processes. Arguing as in the proof of consistency of $\hat{\beta}$, one can show that $\hat{\Gamma}(b) \to \Gamma(b)$ in a neighborhood $\mathcal{B}$ of β, where $\Gamma(b)$ is the derivative of $\mathcal{U}$ and

$$\Gamma(b) = \int_0^{T_0} -\frac{\mathcal{E}\{X^2Ze^{Xb}I(Y \geq s)\}}{\mathcal{E}\{Ze^{Xb}I(Y \geq s)\}} + \frac{\mathcal{E}\{XZe^{Xb}I(Y \geq s)\}^2}{\mathcal{E}\{Ze^{Xb}I(Y \geq s)\}^2} d\mathcal{E}\{\Delta I(Y \leq s)\}$$

Applying Taylor expansion, we have $\hat{U}(\hat{\beta}) - \hat{U}(\beta) = \hat{\Gamma}(\beta^*)(\hat{\beta} - \beta)$, where β^* lies on the segment between $\hat{\beta}$ and β. In light of the consistency of $\hat{\beta}$, and therefore β^*, for β as well as the continuity of $\Gamma(\beta)$ at β, $\hat{\Gamma}(\beta^*)$ converges to $\Gamma(\beta)$ almost surely. By Slutsky's theorem, $\sqrt{n}(\hat{\beta} - \beta)$ converges to a normal distribution with mean zero and covariance matrix $\Gamma(\beta)^{-1}\Sigma(\beta)\{\Gamma(\beta)^{-1}\}^T$, where $\Sigma(\beta) = E[\psi_1(\beta)\psi_1(\beta)^T]$. Arguing as before, $\Gamma(\beta)$ can be consistently estimated by $\hat{\Gamma}(\hat{\beta})$, and, as a result, $\hat{\Gamma}(\hat{\beta})^{-1}\hat{\Sigma}(\hat{\beta})\{\hat{\Gamma}(\hat{\beta})^{-1}\}^T$ is a consistent variance estimator.

Proof of Lemma 3.

Define the functions $\hat{H}_0(t;b) = \int_0^t \frac{d\hat{\mathcal{E}}\{\Delta I(Y \leq s)\}}{\hat{\mathcal{E}}\{\hat{Z}e^{Xb}I(Y \geq s)\}}$ and $H_0(t;b) = \int_0^t \frac{d\mathcal{E}\{\Delta I(Y \leq s)\}}{\mathcal{E}\{Ze^{Xb}I(Y \geq s)\}}$. $\hat{H}_0(t;b)$ is a continuous functional of two processes since the denominator is bounded away from zero. The almost sure convergence of the two processes can be established from the previous discussions. It can be shown that $\sup_{t \in [0,T_0], b \in \mathcal{B}} | \hat{H}_0(t;b) - H_0(t;b) | \to 0$ almost surely. Then the consistency of $\hat{H}_0(t, \hat{\beta})$ for $H_0(t)$ follows the strong consistency of $\hat{\beta}$ for β.

A Taylor expansion of $\hat{H}_0(t;\hat{\beta})$ about β gives

$$\hat{H}_0(t;\hat{\beta}) = \hat{H}_0(t;\beta) + \frac{\partial \hat{H}_0(t;b)}{\partial b}\bigg|_{b=\beta_t^*} (\hat{\beta} - \beta), \tag{6}$$

where β_t^* depends on t and lies on the line segment between $\hat{\beta}$ and β. By similar argument used above, one can show that $\partial\hat{H}_0(t;b)/\partial b \,|_{b=\beta_t^*}$ converges in probability to $\partial H_0(t;b)/\partial b \,|_{b=\beta}$

26

for $t \in [0, T_0]$. Moreover, the functional delta method applied to $\hat{H}_0(t, \beta)$ yields

$$\sqrt{n}(\hat{H}_0(t; \beta) - H_0(t; \beta)) = \int_0^t -\frac{\sqrt{n}(\hat{\mathcal{E}}\{\hat{Z}e^{X\beta}I(Y \geq s)\} - \mathcal{E}\{Ze^{X\beta}I(Y \geq s)\})}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}^2} d\mathcal{E}\{\Delta I(Y \leq s)\}$$

$$+ \int_0^t \frac{d\sqrt{n}(\hat{\mathcal{E}}\{\Delta I(Y \leq s)\} - \mathcal{E}\{\Delta I(Y \leq s)\})}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}} + o_p(1).$$

Following Theorem 2, $\sqrt{n}(\hat{\beta} - \beta) = -\Gamma(\beta)\sqrt{n}\hat{U}(\beta) + o_p(1)$, and by definition $H_0(t, \beta) = H_0(t)$.

From (6) the estimator of the baseline cumulative hazard function can be expressed as

$$\sqrt{n}\left\{\hat{H}_0(t; \hat{\beta}) - H_0(t)\right\} = \sqrt{n}\left\{\hat{H}_0(t; \beta) - H_0(t; \beta)\right\} + \frac{\partial \hat{H}_0(t; b)}{\partial b}\Big|_{b=\beta_t^*} \sqrt{n}(\hat{\beta} - \beta)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i(t) + o_p(1),$$

where $\phi_i(t)$ is defined by

$$\phi_i(t) = \int_0^t -\frac{\psi_i^{(3)}(s; \beta)d\mathcal{E}\{\Delta I(Y \leq s)\}}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}^2}$$

$$+ \int_0^t \frac{d(\Delta_i I(Y_i \leq s) - \mathcal{E}\{\Delta I(Y \leq s)\})}{\mathcal{E}\{Ze^{X\beta}I(Y \geq s)\}} - \frac{\partial H_0(t; \beta)}{\partial \beta}\Gamma(\beta)\psi_i(\beta).$$

Since $\phi_i(t)$ is a linear combination of monotone processes with bounded second moments, the weak convergence of $\sqrt{n}(\hat{H}_0(t; \hat{\beta}) - H_0(t))$ follows from Example 2.11.16 of van der Vaart and Wellner (1996).

REFERENCES

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

Byar, D. P. (1980), "The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine, and Topical Thiotepa.
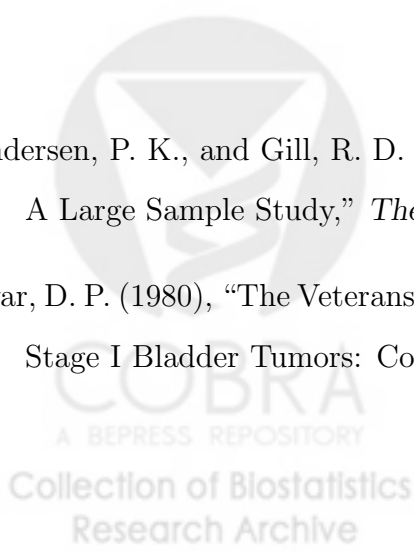
Bladder Tumors and Other Topics in Urological Oncology, New York: Plenum, pp. 363-370.

Chang, S.-H. and Wang, M.-C.(1999), "Conditional Regression Analysis for Recurrence Time Data," *Journal of the American Statistical Association*, 94, 1221–1230.

Clayton, D. G. (1978), "A model for Association in Bivariate Lifetables and Its Application in Epidemiological Studies of Family Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Cox, D. R. (1972), "Regression Models and Life Tables (With Discussion)," *Journal of the Royal Statistical Society B*, 34, 187-220.

Cox, D. R., and Oakes, D. A. (1984), "Analysis of Survival Data," New York: Chapman & Hall.

Eaton, W. W., Mortensen, P. B., Herrman, H., Freeman, H., Bilker, W., Burgess, P., and Wooff, K. (1992a), "Long-term Course of Hospitalization for Schizophrenia: Part I. Risk for Hospitalization," *Schizophrenia Bulletin*, 18, 217–228.

Eaton, W. W., Bilker, W., Haro, J. M., Herrman, H., Mortensen, P. B., Freeman, H., and Burgess P. (1992b), "Long-term Course of Hospitalization for Schizophrenia: Part II. Change With Passage of Time," *Schizophrenia Bulletin*, 18, 229–241.

Ghosh, D., and Lin, D. Y. (2003), "Semiparametric Analysis of Recurrent Events in the Presence of Dependent Censoring," to appear in Biometrics.

Henderson, R., Diggle, P., and Dobson, A. (2000), "Joint Modelling of Longitudinal Measurements and Event Time Data," *Biostatistics*, 1, 465–480.

Huang, Y., and Wang, M.-C. (2002), " Frequency of Recurrent Events at Failure Time: Modeling and Inference," *Journal of the American Statistical Association*, 98, 663–670.

28

Lancaster, T., and Intrator, O. (1998), "Panel Data With Survival: Hospitalization of HIV-positive Patients," *Journal of the American Statistical Association*, 93, 46–53.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000), "Semiparametric Regression for the Mean and Rate Functions of Recurrent Events," *Journal of the Royal Statistical Society B*, 62, 711–730.

Lin, H., Turnbull, B. W., McCulloch, C. E., and Slate, E. H. (2002), "Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data: Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer," *Journal of the American Statistical Association*, 97, 53–65.

Oakes, D. (1989), "Bivariate survival models induced by frailties," *Journal of the American Statistical Association* 84, 487–493.

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

Prentice R. L., Williams, B. J., and Peterson, A. V. (1981), "On the Regression Analysis of Multivariate Failure Time Data," *Biometrika*, 68, 373-379.

Schoenfeld, D. (1982), "Partial Residuals for the Proportional Hazards Regression Model," *Biometrika*, 69, 239–241.

van der Vaart, A. W., and Wellner, J. A. (1996), "Weak Convergence and Empirical Processes," New York: Springer.

Wang, M.-C., Jewell, N. P., and Tsai, W.-Y. (1986), "Asymptotic Properties of the Product Limit Estimate Under Random Truncation," *The Annals of Statistics*, 14, 1597–1650.

29

Wang, M.-C, Qin, J., and Chiang, C.-T. (2001), "Analyzing Recurrent Event Data With
  Informative Censoring," *Journal of the American Statistical Association*, 96, 1057–
  1065.

Wassell, J. T., Wojciechowski W. C., and Landen D. D. (1999), "Recurrent Injury Event-
  time Analysis," *Statistics in Medicine*, 18, 3355–3363.

30

Table 1: *Summary of the simulated data*

(α,β)	$Z \sim poisson(10)$			$Z \sim gamma(2,5)$		
	$P(death)$	Y	m	$P(death)$	Y	m
Scenario I: $\lambda_0(t) = 1/10$, $h_0(t) = t/400$						
$(0,0)$	0.27	3.90	3.64	0.26	4.44	3.40
$(-1,-1.5)$	0.14	4.41	2.34	0.14	4.91	2.19
Scenario II: $\lambda_0(t) = (t+1)/60$, $h_0(t) = \sqrt{t}/200$						
$(0,0)$	0.28	3.80	2.40	0.26	4.35	2.26
$(-1,-1.5)$	0.15	4.36	1.59	0.14	4.87	1.58

$P(death)$ is the average death rate; Y is the average terminating time; m is the average number of recurrent events

Table 2: *Summary statistics of the simulation studies*

(α,β)	$Z \sim poisson(10)$							$Z \sim gamma(2,5)$						
	B_α	V_α	B_β	V_β	ρ	$Avg\ \hat\beta^*$	$V_{\beta*}$	B_α	V_α	B_β	V_β	ρ	$Avg\ \hat\beta^*$	$V_{\beta*}$
Scenario I: $\lambda_0(t) = 1/10$, $h_0(t) = t/400$														
$n=200$														
$(0,0)$	0	181	-3	383	0.43	0.383	334	4	274	-4	438	0.61	0.584	331
$(-1,-1.5)$	-8	255	-25	487	0.49	-1.228	406	-5	277	-25	484	0.50	-0.769	395
$n=500$														
$(0,0)$	-3	151	-1	243	0.58	0.244	191	7	199	21	283	0.67	0.599	207
$(-1,-1.5)$	0	176	-6	307	0.54	-1.211	249	-13	212	-1	324	0.63	-0.752	244
Scenario II: $\lambda_0(t) = (t+1)/60$, $h_0(t) = \sqrt{t}/200$														
$n=200$														
$(0,0)$	-13	412	2	496	0.75	0.211	305	17	511	23	608	0.83	0.539	319
$(-1,-1.5)$	-38	422	-44	558	0.64	-1.253	400	-51	553	-17	681	0.78	-0.846	410
$n=500$														
$(0,0)$	10	243	15	303	0.72	0.208	197	4	358	16	399	0.86	0.538	192
$(-1,-1.5)$	-23	287	-16	373	0.71	-1.246	246	-18	369	-15	445	0.81	-0.822	246

B_α and B_β are the empirical bias ($\times 1000$) of $\hat\alpha$ and $\hat\beta$; V_α and V_β are the empirical standard error ($\times 1000$) of $\hat\alpha$ and $\hat\beta$; ρ is the empirical correlation coefficient of $\hat\alpha$ and $\hat\beta$; $Avg\ \hat\beta^*$ is the empirical average and $V_{\beta*}$ is the empirical standard error ($\times 1000$) of the estimator based on the Cox proportional hazards model.
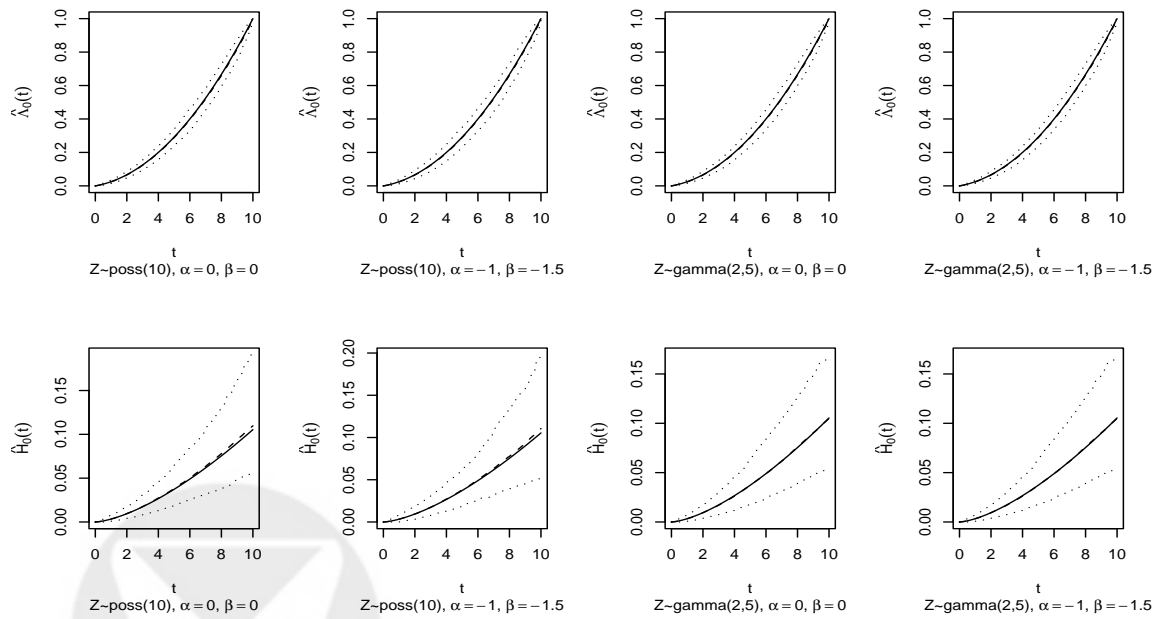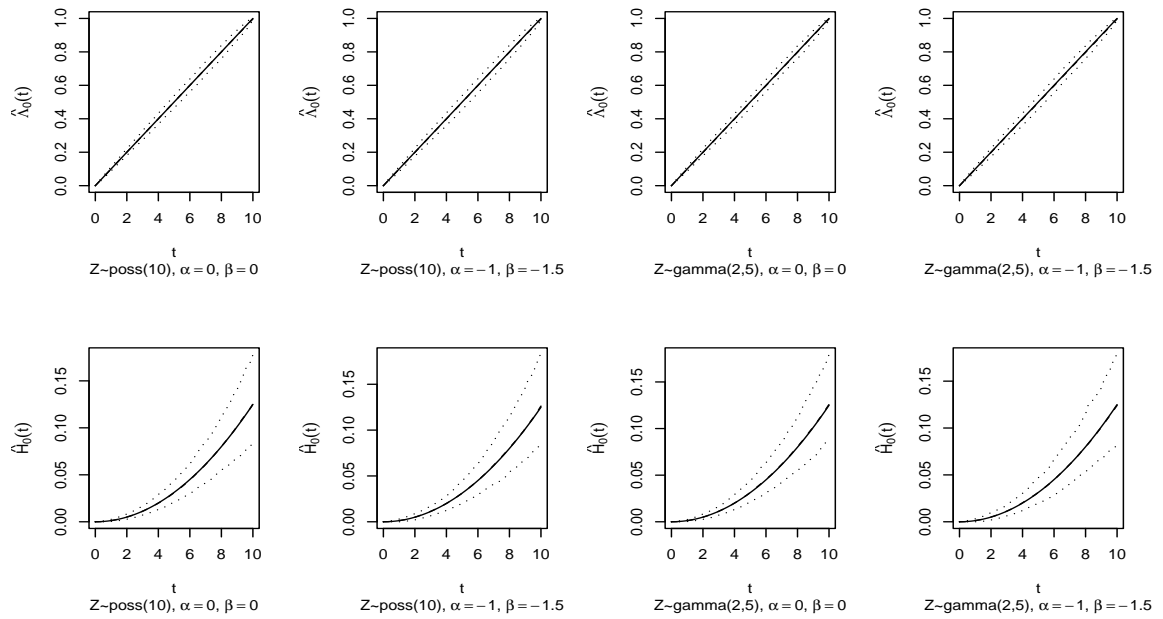
31
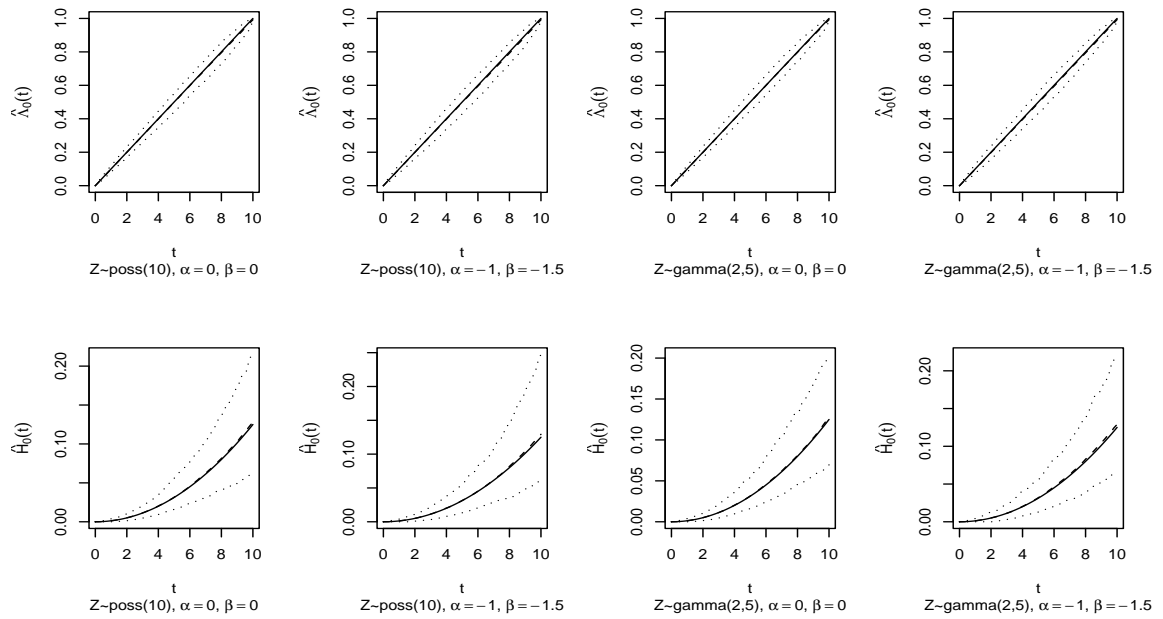
Figure 1: *Plots of estimated $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ with pointwise 95% confidence intervals for $n = 200$. Scenario I: $\lambda_0(t) = 1/10$, and $h_0(t) = t/400$; Scenario II:$\lambda_0(t) = (t+1)/60$, $h_0(t) = \sqrt{t}/200$. —: True curve, - - - : empirical average, $\cdots$: pointwise 95% confidence intervals.*
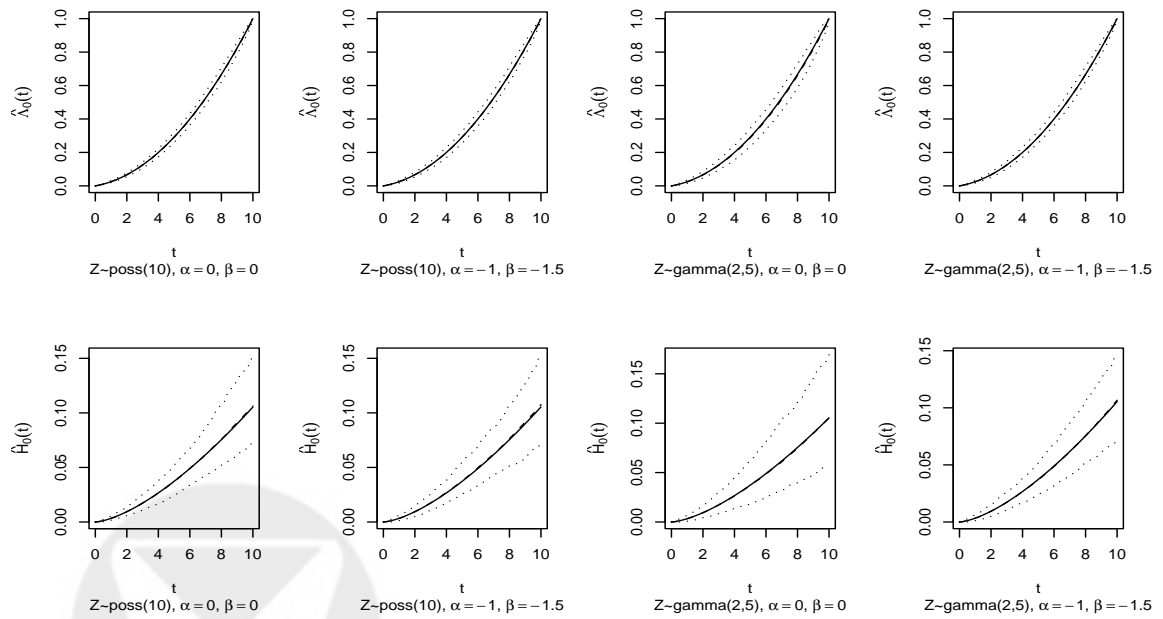
32

Scenario I

Scenario II

Figure 2: *Plots of estimated $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ with pointwise 95% confidence intervals for $n = 500$. Scenario I: $\lambda_0(t) = 1/10$, and $h_0(t) = t/400$; Scenario II: $\lambda_0(t) = (t+1)/60$, $h_0(t) = \sqrt{t}/200$.—: True curve, - - - : empirical average, $\cdots$: pointwise 95% confidence intervals.*

33

Table 3: *Hospital admissions and deaths for different subgroups*

Subgroup	No. patients	No. deaths	No. of hospital admissions since entry						
			0	1	2	3	4	5	≥6
Male	3318	368	984	581	394	331	200	157	671
(%)	100	11.1	29.7	17.5	11.9	10	6	4.7	20.2
Female	5493	685	1392	945	636	470	363	279	1408
(%)	100	12.5	25.3	17.2	11.6	8.6	6..6	5.1	25.6
Onset Age≤20	1065	76	187	130	144	90	82	59	373
(%)	100	7.1	17.6	12.2	13.5	8.5	7.7	5.5	35.0
Onset Age>20	7746	977	2189	1396	886	711	481	377	1706
(%)	100	12.6	28.3	18.0	11.4	9.2	6.2	4.9	22.0

Table 4: *Summary of Denmark PCR data analysis*

Risk factor	Estimate	SE	95% bootstrap CI
Hospital admissions:			
Onset Age≤ 20	0.19	0.04	(0.10, 0.27)
Gender	-0.18	0.04	(-0.26, -0.09)
Death:			
Onset Age≤ 20	-0.84	0.13	(-1.10, -0.62)
Gender	-0.11	0.07	(-0.25, 0.01)

Note: SE= standard error of estimates from the 200 bootstrap samples;
95% bootstrap CI= (2.5%, 97.5%) quantiles of the 200 estimates.

Baseline Cumulative Rate Function
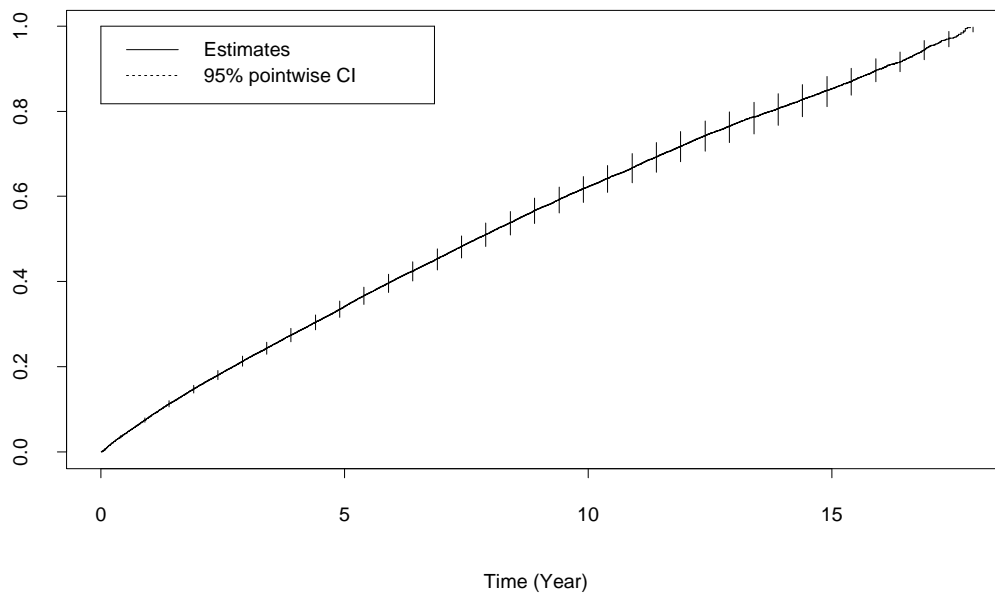
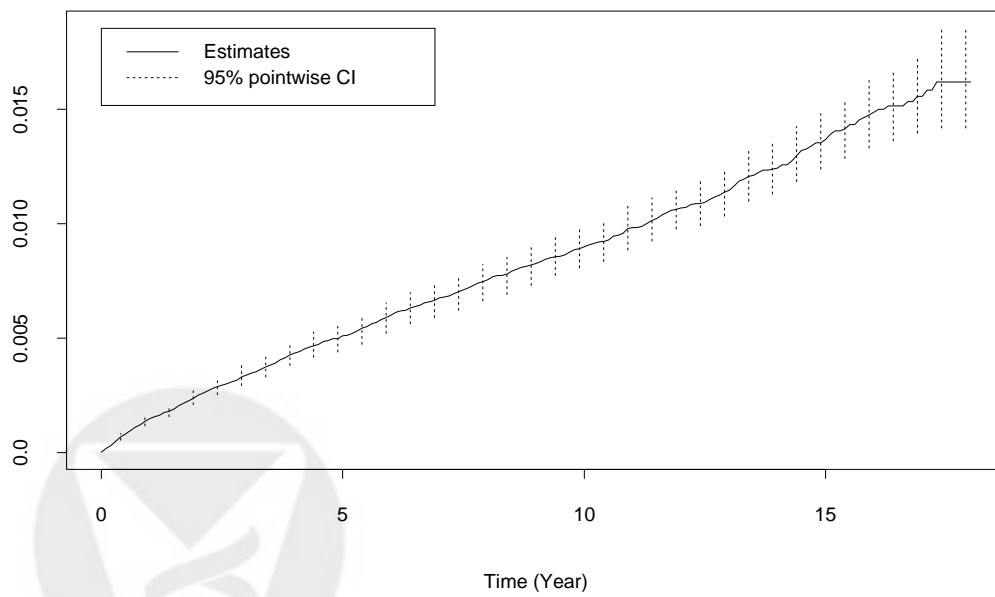Baseline Cumulative Hazard Function

Figure 3: *Plots of $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ for the Denmark schizophrenia cohort data, with pointwise 95% bootstrap confidence intervals.*

35