Supplementary Material to "Checking semiparametric transformation models with censored data"

LI CHEN

Markey Cancer Center and Department of Biostatistics, University of Kentucky, Lexington, KY 40536, USA

D. Y. LIN AND DONGLIN ZENG

Department of Biostatistics, CB 7420, University of North Carolina, Chapel Hill, NC 27599, USA

S.1. Weak convergence of $W_n(x,t)$

We establish the weak convergence of $W_n(x,t)$ for general f_1 and f_2 that satisfy the following conditions.

Condition 1. There exists some $\delta > 0$ such that $\{f_1(\overline{X}(t);\beta,\Lambda); t \in [0,\tau], \beta \in \mathcal{U}(\beta_0), \Lambda \in \mathcal{U}(\Lambda_0)\}$ is a uniformly bounded P_0-Donsker class and $\{f_2(\overline{X}(t);\beta,\Lambda); t \in [0,\tau], \beta \in \mathcal{U}(\beta_0), \Lambda \in \mathcal{U}(\Lambda_0)\}$ is a P_0-Donsker class, where $\mathcal{U}(\beta_0) = \{\beta : |\beta - \beta_0| < \delta\}$ and $\mathcal{U}(\Lambda_0) = \{\Lambda : \sup_{t\in[0,\tau]} |\Lambda(t) - \Lambda_0(t)| < \delta\}$.

Condition 2. With probability one, there exist constants K_1 and K_2 such that $f_1(\overline{X}(t);\beta_0, \Lambda_0)$ has total variation bounded by K_1, and $I(f_2(\overline{X}(t);\beta_0,\Lambda_0) \le x)$ as a process indexed by t has total variation bounded by K_2 for all x.

Condition 3. $\sup_t E|f_1(\overline{X}(t);\beta_n,\Lambda_n) - f_1(\overline{X}(t);\beta_0,\Lambda_0)| \to 0$ and $\sup_{x,t} E|I(f_2(\overline{X}(t);\beta_n, \Lambda_n) \le x) - I(f_2(\overline{X}(t);\beta_0,\Lambda_0) \le x)| \to 0$ as $\beta_n \to \beta_0$ and $\Lambda_n \to \Lambda_0$ in $l^\infty[0,\tau]$.

Remark S.1. Conditions 1 and 2, which ensure the uniform weak convergence, are satisfied by all the processes considered in the main paper. Condition 3 pertains to continuity requirements on the functions f_1 and f_2. The continuity condition on f_1 is satisfied by all the processes of the main paper. Since f_2 does not involve $\widehat{\beta}$ or $\widehat{\Lambda}$ in the processes W_o, W_c

and W_p, the continuity condition on f_2 is automatically satisfied. It can be shown that this condition is satisfied for the processes W_l and W_{tr} if there is at least one continuous covariate. In addition, the condition of at least one continuous covariate is necessary for W_l and W_{tr}. Suppose, for example, that X in W_l consists of a single discrete covariate and $\beta_0 > 0$. Then in Condition 3, $E|I(f_2(\overline{X}(t);\beta_n,\Lambda_n) \le x) - I(f_2(\overline{X}(t);\beta_0,\Lambda_0) \le x)| = E|I(X \le x/\beta_n) - I(X \le x/\beta_0)| \ge |E\{I(X \le x/\beta_n)\} - E\{I(X \le x/\beta_0)\}|$, which may not converge to 0 when x is the mass point of $\beta_0 X$.

Define

$$\begin{aligned} g &= \int_0^t f(\overline{X}(u);x,\beta_0,\Lambda_0)dM(u;\beta_0,\Lambda_0), \\ \widehat{g}_n &= \int_0^t f(\overline{X}(u);x,\widehat{\beta},\widehat{\Lambda})dM(u;\widehat{\beta},\widehat{\Lambda}), \\ g_{1n} &= \int_0^t f(\overline{X}(u);x,\beta_0,\Lambda_0)dM(u;\widehat{\beta},\widehat{\Lambda}), \\ g_{2n} &= \int_0^t f(\overline{X}(u);x,\widehat{\beta},\widehat{\Lambda})dM(u;\beta_0,\Lambda_0). \end{aligned}$$

We first prove two lemmas which will be used in the proof of the weak convergence of $W_n(x,t)$.

Lemma S.1. $\int(\widehat{g}_n - g)^2 dP_0 \xrightarrow{p} 0$ uniformly for $x \in \mathcal{R}^q$ and $t \in [0,\tau]$.

Proof. For any $x \in \mathcal{R}^q$ and $t \in [0,\tau]$,

$$\begin{aligned} & E|\widehat{g}_n - g| \\ = \ & E\left|\int_0^t f(\overline{X}(u);x,\widehat{\beta},\widehat{\Lambda})dM(u;\widehat{\beta},\widehat{\Lambda}) - \int_0^t f(\overline{X}(u);x,\beta_0,\Lambda_0)dM(u;\beta_0,\Lambda_0)\right| \\ \le \ & \int_0^t [E|f(\overline{X}(u);x,\widehat{\beta},\widehat{\Lambda}) - f(\overline{X}(u);x,\beta_0,\Lambda_0)|^2]^{1/2}[E|dM(u;\widehat{\beta},\widehat{\Lambda})|^2]^{1/2} \\ & + E\left|\frac{1}{\sqrt{n}}\int_0^t f(\overline{X}(u);x,\beta_0,\Lambda_0)Y(u)d\{\sqrt{n}(\widetilde{G}(u;\widehat{\beta},\widehat{\Lambda}) - \widetilde{G}(u;\beta_0,\Lambda_0))\}\right|. \end{aligned} \tag{S.1}$$

Under Conditions 2 and 3, $\sup_{x,u}[E|f(\overline{X}(u);x,\widehat{\beta},\widehat{\Lambda}) - f(\overline{X}(u);x,\beta_0,\Lambda_0)|^2]^{1/2} \xrightarrow{p} 0$. In addition,

$$E|dM(u;\widehat{\beta},\widehat{\Lambda})|^2 \le K\left(E[dN(u)] + E\left[\left\{G'(\int_0^u e^{\widehat{\beta}^{\mathrm{T}}X(s)}d\widehat{\Lambda}(s))e^{\widehat{\beta}^{\mathrm{T}}X(s)}\right\}^2\right](d\widehat{\Lambda}(u))^2\right)$$

for some constant K. Thus, the first term on the right-hand side of the inequality in (S.1) converges uniformly to 0 in probability.

As shown in the above proof, $\widetilde{G}(u;\beta,\Lambda)$ is Hadamard differentiable with respect to (β,Λ). By the functional delta-method, $\sqrt{n}\{\widetilde{G}(u;\widehat{\beta},\widehat{\Lambda})-\widetilde{G}(u;\beta_0,\Lambda_0)\}$ converges weakly to a zero-mean Gaussian process in the metric space $l^\infty(\mathcal{F})$ for almost every path of $X(\cdot)$. It can be verified that there exists a constant K such that $f(\overline{X}(u);x,\beta_0,\Lambda_0)$ has total variation bounded by K for all x. Thus, the second term on the right-hand side of the inequality in (S.1) converge uniformly to 0 in probability.

It is easy to show that $\widehat{g}_n$ and g are uniformly bounded by $K(N(\tau)+1)$ in probability for some constant K. Thus,

$$\begin{aligned} E|\widehat{g}_n-g|^2 &\le KE[(N(\tau)+1)|\widehat{g}_n-g|] \\ &\le 2K^2E[I(N(\tau)>M)(N(\tau)+1)^2]+K(M+1)E[I(N(\tau)\le M)|\widehat{g}_n-g|] \end{aligned}$$

for every $M>0$. Since $E|\widehat{g}_n-g|$ converges uniformly to 0 in probability,

$$\limsup_n \sup_{x,t} E|\widehat{g}_n-g|^2 \le 2K^2E[I(N(\tau)>M)(N(\tau)+1)^2].$$

We then obtain the result of Lemma S.1 by letting M go to ∞.

Lemma S.2. Let $A_n(u,x,t)$ be stochastic processes with $\sup_{u,x,t}|A_n(u,x,t)| \xrightarrow{p} 0$. Then $\sqrt{n}\int_0^\tau A_n(u,x,t)d(\widehat{\Lambda}-\Lambda_0)(u)$ converges uniformly to 0 in probability.

Proof. Clearly,

$$\sup_{x,t}\left|\sqrt{n}\int_0^\tau A_n(u,x,t)d(\widehat{\Lambda}-\Lambda_0)(u)\right| \le \sup_{u,x,t}|A_n(u,x,t)|\int_0^\tau |\sqrt{n}d(\widehat{\Lambda}-\Lambda_0)(u)|. \qquad \text{(S.2)}$$

Note that $\sup_{u,x,t}|A_n(u,x,t)| \to_p 0$. In addition, $\sqrt{n}(\widehat{\Lambda}-\Lambda_0)$ converges weakly to a zero-mean Gaussian process with bounded total variation (Zeng and Lin 2006). By Slutsky's lemma and Skorohod's representation theorem, the right-hand side of the inequality in (S.2) converges to 0 in probability. We then obtain the result of the lemma.

Now we prove the weak convergence of $W_n(x,t)$. Because $P_0 g_{2n} = o_p(n^{-1/2})$, $W_n(x,t)$ can be written as

$$\sqrt{n}(\mathcal{P}_n - P_0)(\widehat{g}_n - g) + \sqrt{n}(\mathcal{P}_n - P_0)g + \sqrt{n}P_0(g_{1n} - g) + \sqrt{n}P_0\{(\widehat{g}_n - g_{2n}) - (g_{1n} - g)\}. \quad \text{(S.3)}$$

The first term in (S.3) converges uniformly to 0 in probability because, as stated in Lemma S.1, $\int(\widehat{g}_n - g)^2 dP_0 \xrightarrow{p} 0$ uniformly in x and t and because $\widehat{g}_n$ and g belong to a P_0-Donsker class under Condition 1 and the conditions on $X(\cdot)$ stated in Section 2 of the main paper, which is justified as follows. By Lemma 4.1 of Kosorok (2008), $\{X(t); t \geq 0\}$ is a P_0-Donsker class. It is also clear that $\{N(t); t \geq 0\}$, $\{Y(t); t \geq 0\}$ and $\{\Lambda(t); t \geq 0\}$ are P_0-Donsker classes. Because $\int_0^t Y(s)e^{\widehat{\beta}^{\mathrm{T}}X(s)}d\widehat{\Lambda}(s)$ belongs to $\mathcal{F}^{(P,2)}$, where $\mathcal{F}^{(P,2)}$ denotes the pointwise and $L_2(P)$ closure of the class $\big\{\sum_{i=1}^{m} I(t_i \leq t)Y(t_i)e^{\beta^{\mathrm{T}}X(t_i)}(\Lambda(t_i) - \Lambda(t_{i-1})); t \in [0,\tau], m > 0, 0 = t_0 < t_1 < \ldots < t_m = \tau, \beta \in \mathcal{U}(\beta_0), \Lambda \in \mathcal{U}(\Lambda_0)\big\}$, the function $\int_0^t Y(s)e^{\widehat{\beta}^{\mathrm{T}}X(s)}d\widehat{\Lambda}(s)$ belongs to a P_0-Donsker class by the preservation properties of P_0-Donsker class (van der Vaart and Wellner 1996, ch2.10). Since G has continuous first derivatives, $G(\int_0^t Y(s)e^{\widehat{\beta}^{\mathrm{T}}X(s)}d\widehat{\Lambda}(s))$ belongs to a P_0-Donsker class. Therefore, $M(t, \widehat{\beta}, \widehat{\Lambda})$ belongs to a P_0-Donsker class and has uniformly bounded total variation. By the same arguments of the pointwise closure, we then conclude that $\widehat{g}_n$ belongs to a P_0-Donsker class.

The second term in (S.3) converges weakly to a zero-mean Gaussian process in $l^\infty(\mathcal{R}^q \times [0,\tau])$ since g also belongs to a P_0-Donsker class.

To study the last two terms in (S.3), we define $\widetilde{G}(t;\beta,\Lambda) = G\big\{\int_0^t e^{\beta^{\mathrm{T}}X(s)}d\Lambda(s)\big\}$. Then the third and fourth terms in (S.3) can be written as

$$\sqrt{n}P_0 \int_0^t f(\overline{X}(u); x, \beta_0, \Lambda_0)Y(u)\{d\widetilde{G}(u;\widehat{\beta},\widehat{\Lambda}) - d\widetilde{G}(u;\beta_0,\Lambda_0)\}, \quad \text{(S.4)}$$

and

$$\sqrt{n}P_0 \int_0^t \{f(\overline{X}(u); x, \widehat{\beta}, \widehat{\Lambda}) - f(\overline{X}(u); x, \beta_0, \Lambda_0)\}Y(u)\{d\widetilde{G}(u;\widehat{\beta},\widehat{\Lambda}) - d\widetilde{G}(u;\beta_0,\Lambda_0)\}, \quad \text{(S.5)}$$

respectively. It can be verified that $\int_0^t e^{\beta^{\mathrm{T}}X(s)}d\Lambda(s)$ is Hadamard differentiable with respect to (β, Λ) for almost every path of $X(\cdot)$. By the chain rule, $\widetilde{G}(t;\beta,\Lambda)$ is Hadamard differentiable. Then the weak convergence of $\sqrt{n}(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0)$ (Zeng and Lin 2006) implies that $\sqrt{n}\{\widetilde{G}(t;\widehat{\beta},\widehat{\Lambda})-\widetilde{G}(t;\beta_0,\Lambda_0)\}-\sqrt{n}\widetilde{G}'_{\beta_0,\Lambda_0}(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0)(t)$ converges to 0 in probability uniformly in $t\in[0,\tau]$, where $\widetilde{G}'_{\beta_0,\Lambda_0}$ is the Hadamard derivative of $\widetilde{G}(\cdot;\beta,\Lambda)$ at (β_0,Λ_0). It follows from Conditions 1-3 and integration by part that $\sqrt{n}P_0\int_0^t f(\overline{X}(u);x,\beta,\Lambda)Y(u)\{d\widetilde{G}(u;\widehat{\beta},\widehat{\Lambda})-d\widetilde{G}(u;\beta_0,\Lambda_0)\}$ is asymptotically equivalent to

$$\begin{aligned}&\sqrt{n}P_0\int_0^t f(\overline{X}(u);x,\beta,\Lambda)Y(u)d\{\widetilde{G}'_{\beta_0,\Lambda_0}(\widehat{\beta}-\beta_0,\widehat{\Lambda}-\Lambda_0)\}\\ =\;&\sqrt{n}(\widehat{\beta}-\beta_0)^{\mathrm{T}}P_0h_1(Y,X;x,t;\beta,\Lambda)+\sqrt{n}\int_0^\tau P_0h_2(Y,X;u,x,t;\beta,\Lambda)d(\widehat{\Lambda}-\Lambda_0)(u)\end{aligned}$$

uniformly in $x\in\mathcal{R}^q$, $t\in[0,\tau]$, $\beta\in\mathcal{U}(\beta_0)$ and $\Lambda\in\mathcal{U}(\Lambda_0)$. Thus, (S.4) and (S.5) are asymptotically equivalent to

$$\sqrt{n}(\widehat{\beta}-\beta_0)^{\mathrm{T}}P_0h_1(Y,X;x,t;\beta_0,\Lambda_0)+\sqrt{n}\int_0^\tau P_0h_2(Y,X;u,x,t;\beta_0,\Lambda_0)d(\widehat{\Lambda}-\Lambda_0)(u), \tag{S.6}$$

and

$$\begin{aligned}&\sqrt{n}(\widehat{\beta}-\beta_0)^{\mathrm{T}}P_0(h_1(Y,X;x,t;\widehat{\beta},\widehat{\Lambda})-h_1(Y,X;x,t;\beta_0,\Lambda_0))\\ &+\sqrt{n}\int_0^\tau P_0(h_2(Y,X;u,x,t;\widehat{\beta},\widehat{\Lambda})-h_2(Y,X;u,x,t;\beta_0,\Lambda_0))d(\widehat{\Lambda}-\Lambda_0)(u),\end{aligned}\tag{S.7}$$

respectively. Because $\sqrt{n}((\widehat{\beta}-\beta_0)^{\mathrm{T}}\widetilde{h}_1+\int_0^\tau\widetilde{h}_2(u)d(\widehat{\Lambda}-\Lambda_0)(u))$ is asymptotically equivalent to $\sqrt{n}(\mathcal{P}_n-P_0)(S_{\beta_0},S_{\Lambda_0})I^{-1}_{\beta_0,\Lambda_0}(\widetilde{h}_1,\widetilde{h}_2(\cdot))$ for all $(\widetilde{h}_1,\widetilde{h}_2(\cdot))\in\mathcal{R}^p\times\mathcal{F}$, where $\mathcal{F}=\{w(t):\|w\|_{BV[0,\tau]}\le 1\}$ and $\|w\|_{BV[0,\tau]}$ denotes the total variation of $w(\cdot)$ in $[0,\tau]$ (Zeng and Lin 2006), (S.6) is further asymptotically equivalent to

$$\sqrt{n}(\mathcal{P}_n-P_0)(S_{\beta_0},S_{\Lambda_0})I^{-1}_{\beta_0,\Lambda_0}(P_0h_1(Y,X;x,t;\beta_0,\Lambda_0),P_0h_2(Y,X;\cdot,x,t;\beta_0,\Lambda_0)).$$

Under Condition 3,

$$\sup_{x,t}|P_0(h_1(Y,X;x,t;\widehat{\beta},\widehat{\Lambda})-h_1(Y,X;x,t;\beta_0,\Lambda_0))|\xrightarrow{p}0,$$

$$\sup_{u,x,t}|P_0(h_2(Y,X;u,x,t;\widehat{\beta},\widehat{\Lambda})-h_2(Y,X;u,x,t;\beta_0,\Lambda_0))|\xrightarrow{p}0.$$

Hence, (S.7) converges uniformly to zero by Lemma S.2.

Combining the above results, we obtain the weak convergence of $W_n(x,t)$.

S.2. Validity of the Monte Carlo Procedure

In this section, we show that the conditional distribution of $\widehat{W}_n(x,t)$ given the observed data has the same limiting distribution as $\widetilde{W}_n(x,t)$. This result will follow upon verifying the conditions in Theorem 2.11.1 of van der Vaart and Wellner (1996). In this case, the functional class $\mathcal{F}$ is indexed by x and t, and the semimetric ρ is defined as the Euclidean metric in (x,t).

Define

$$
\begin{aligned}
\psi_i(x,t) = & \int_0^t f_1(\overline{X}_i(u);\beta_0,\Lambda_0) I(f_2(\overline{X}_i(u);\beta_0,\Lambda_0) \le x) dM_i(u;\beta_0,\Lambda_0) \\
& + (S_{\beta_0}, S_{\Lambda_0})_i I^{-1}_{\beta_0,\Lambda_0}(P_0 h_1(Y,X;x,t;\beta_0,\Lambda_0), P_0 h_2(Y,X;\cdot,x,t;\beta_0,\Lambda_0)), \\
\widehat{\psi}_i(x,t) = & \int_0^t f_1(\overline{X}_i(u);\widehat{\beta},\widehat{\Lambda}) I(f_2(\overline{X}_i(u);\widehat{\beta},\widehat{\Lambda}) \le x) dM_i(u;\widehat{\beta},\widehat{\Lambda}) + S_i(x,t),
\end{aligned}
$$

where $(S_{\beta_0}, S_{\Lambda_0})_i$ is the score operator for β and Λ at (β_0,Λ_0) from the ith subject, and $S_i(x,t) = l_i^{\mathrm{T}} I_n^{-1} (h_{1n}^{\mathrm{T}}(x,t), h_{2n}^{\mathrm{T}}(x,t))^{\mathrm{T}}$.

Condition 1 in Theorem 2.11.1 of van der Vaart and Wellner (1996) requires that

$$
\frac{1}{n}\sum_{i=1}^n \sup_{x,t\in[0,\tau]} |\widehat{\psi}_i(x,t)|^2 E\left[Q_i^2 \left\{ |Q_i| > \frac{\sqrt{n}\eta}{\sup_{x,t\in[0,\tau]} |\widehat{\psi}_i(x,t)|} \right\} \right] \to_p 0 \quad \text{for every } \eta > 0.
$$

This holds since $\sup_{x,t\in[0,\tau]} |\widehat{\psi}_i(x,t)|$ is bounded in probability.

Condition 2 in Theorem 2.11.1 of van der Vaart and Wellner (1996) requires that

$$
\sup_{\rho((x_1,t_1),(x_2,t_2))<\delta_n} \frac{1}{n}\sum_{i=1}^n (\widehat{\psi}_i(x_1,t_1) - \widehat{\psi}_i(x_2,t_2))^2 \to 0 \quad \text{for every } \delta_n \to 0.
$$

To check this condition, note that the first and second terms of $\widehat{\psi}_i(x,t)$ converge in probability to the corresponding first and second terms of $\psi_i(x,t)$. In addition, the two terms and their squares belong to a Glivenko-Cantelli class. Thus, the left-hand side converges in probability to

$$
\lim_{\delta_n \to 0} \sup \sup_{\rho((x_1,t_1),(x_2,t_2))<\delta_n} E[(\psi_i(x_1,t_1) - \psi_i(x_2,t_2))^2],
$$

which can be shown to be zero.

By Theorems 2.6.7 and 2.7.11 of van der Vaart and Wellner (1996), Condition 3 in Theorem 2.11.1 holds since $\mathcal{F}$ indexed by x is a VC-class and $\mathcal{F}$ indexed by t is a class that is Lipschitz in parameter.

Finally, we show that the sequence of the covariance functions converges point-wise to the same limit as that of $\widetilde{W}_n(x,t)$. By the arguments similar to that for checking Condition 2, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}_i(x_1,t_1)\widehat{\psi}_i(x_2,t_2)\longrightarrow_p E\{\psi_i(x_1,t_1)\psi_i(x_2,t_2)\}.$$

This verifies the pointwise convergence of the covariance functions.

S.3. Consistency of Supremum Tests

The proofs rely on the convergence of $\widehat{\beta}$ and $\widehat{\Lambda}$ under misspecified models, which is given in Lemma S.3.

Lemma S.3. Assume that Conditions 2-4 of Zeng and Lin (2006) hold. Let $p_{\beta,\Lambda}$ and p_0 denote the densities of $\{N(t), Y(t), X(t); t \in [0,\tau]\}$ under the posited transformation model (1.2) and under the true model, respectively. Assume that the Kullback-Leibler information between them, i.e., $P_0\log(p_{\beta,\Lambda}/p_0)$, has a unique maximizer (β^*,Λ^*) with Λ^* continuously differentiable in $[0,\tau]$ and (β^*,Λ^*) in the interior of the parameter space. Then the NPMLE $(\widehat{\beta},\widehat{\Lambda})$ under posited transformation model converges to (β^*,Λ^*).

Proof. Using the arguments in Steps 1-2 of Zeng and Lin (2006), we obtain that for every subsequence, there exists a further subsequence such that $\widehat{\beta}\to\widetilde{\beta}$ and $\widehat{\Lambda}\to\widetilde{\Lambda}$. The theorem is proved if we can show that $\widetilde{\beta}=\beta^*$ and $\widetilde{\Lambda}(t)=\Lambda^*(t)$ for $t\in[0,\tau]$.

Define

$$\Lambda_n^*(t)=n^{-1}\int_0^t\sum_{i=1}^{n}\frac{dN_i(s)}{|\phi_n(s;\Lambda^*,\beta^*)|},$$

where ϕ_n is defined in Zeng and Lin (2006). Since it is the maximizer of $P_0\log p_{\beta,\Lambda}$ and lies in the interior of the parameter space, (β^*, Λ^*) is a solution to the score equations and thus

$$\Lambda^*(t) = \int_0^t \frac{EdN(s)}{\phi(s; \Lambda^*, \beta^*)}, \tag{S.8}$$

where

$$\phi(t; \Lambda, \beta) = E\left\{Y(t)e^{\beta^{\mathrm{T}}X(t)}\left[G'\left(\int_0^\tau Y(s)e^{\beta^{\mathrm{T}}X(s)}d\Lambda\right) - \int_t^\tau \frac{G''}{G'}\left\{\int_0^s Y(u)e^{\beta^{\mathrm{T}}X(u)}d\Lambda\right\}dN(s)\right]\right\}.$$

By the Glivenko-Cantelli theorem, Λ_n^* converges to Λ^* uniformly.

Clearly, $\mathcal{P}_n(\log p_{\widehat{\beta},\widehat{\Lambda}} - \log p_{\beta^*,\Lambda_n^*}) \geq 0$. Taking limits on both sides and using the arguments of Zeng and Lin (2006), we obtain $P_0 \log p_{\widetilde{\beta},\widetilde{\Lambda}} \geq P_0 \log p_{\beta^*,\Lambda^*}$. This result, together with the condition that (β^*, Λ^*) is the unique maximizer of $P_0 \log(p_{\beta,\Lambda}/p_0)$, implies that $\beta^* = \widetilde{\beta}$ and $\Lambda^* = \widetilde{\Lambda}$. The convergence of $\widehat{\Lambda}(t) \to \Lambda^*(t)$ can be strengthened to uniform convergence in $t \in [0, \tau]$ by the continuity of Λ^*.

S.3.1. Omnibus test. It suffices to prove that the limit of $\sup_{x,t}|n^{-1/2}W_o(x,t)|$ is nonzero under the alternative hypothesis. By Lemma S.3, $\widehat{\beta} \to \beta^*$ and $\widehat{\Lambda}(t) \to \Lambda^*(t)$. Thus, $n^{-1/2}W_o(x,t)$ converges almost surely to

$$E\left[\int_0^t I(X \leq x)Y(u)\{d\Lambda(u|X) - dG(e^{\beta^{*T}X}\Lambda^*(u))\}\right],$$

which will be nonzero for some $t > 0$ and x under the alternative.

S.3.2. Functional forms of covariates. By Lemma S.3, $\widehat{\beta} \to \beta^*$, $\widehat{\gamma} \to \gamma^*$ and $\widehat{\Lambda}(t) \to \Lambda^*(t)$. Thus, $n^{-1/2}W_c^{(j)}(x,t)$ converges almost surely to

$$E\left[\int_0^t I(X^{(j)} \leq x)Y(u)\{dG(\Lambda_0(u)e^{\beta_0^{\mathrm{T}}X^{(-j)}}g(X^{(j)})) - dG(\Lambda^*(u)e^{\beta^{*T}X^{(-j)}+\gamma^*X^{(j)}})\}\right], \tag{S.9}$$

where $X^{(j)}$ is the jth component of X, and $X^{(-j)}$ consists of the other components of X.

Suppose that the limit of $\sup_{x,t}|n^{-1/2}W_c^{(j)}(x,t)|$ is zero under the alternative. Then (S.9) equals zero for all x and t. Taking the derivative of (S.9) with respect to t at $t = 0$, we have

$$E\left[I(X^{(j)} \leq x)\{G'(0)\lambda_0(0)e^{\beta_0^{\mathrm{T}}X^{(-j)}}g(X^{(j)}) - G'(0)\lambda^*(0)e^{\beta^{*T}X^{(-j)}+\gamma^*X^{(j)}}\}\right] = 0,$$

where λ_0 and λ^* are the derivatives of Λ_0 and Λ^*, respectively. Since $X^{(-j)}$ and $X^{(j)}$ are independent,

$$g(x) = \frac{\lambda^*(0)}{\lambda_0(0)} \frac{Ee^{\beta^{*T} X^{(-j)}}}{Ee^{\beta_0^{\mathrm{T}} X^{(-j)}}} e^{\gamma^* x}.$$

Thus, $\log g(x)$ is linear in x, which is a contradiction.

S.3.3. Link function. The limit of $n^{-1/2}W_l(x,t)$ under the alternative is

$$E\left[\int_0^t I(\beta^{*T}X \le x)Y(u)\{dG(\Lambda_0(u)g(e^{\beta_0^{\mathrm{T}}X})) - dG(\Lambda^*(u)e^{\beta^{*T}X})\}\right]. \tag{S.10}$$

Suppose that the limit of $\sup_{x,t}|n^{-1/2}W_l(x,t)|$ is zero under the alternative. Then (S.10) equals zero for all x and t. Taking the derivative of (S.10) with respect to t at $t=0$, we obtain $E\big[I(\beta^{*T}X \le x)\{G'(0)\lambda_0(0)g(e^{\beta_0^{\mathrm{T}}X}) - G'(0)\lambda^*(0)e^{\beta^{*T}X})\}\big] = 0$. Thus, $E\left[g(e^{\beta_0^{\mathrm{T}}X})|e^{\beta^{*T}X} = x\right] = \lambda^*(0)/\lambda_0(0)x$. By the condition of the theorem, $g(x) = cx^\alpha$ for some constants c and α, which is a contradiction.

S.3.4. Proportionality. Under the alternative, the limit of $n^{-1/2}W_p(t)$ is

$$E\int_0^t Y(u)X\left\{1 + \frac{G''\left(e^{\beta^*X}\Lambda^*(u)\right)}{G'\left(e^{\beta^*X}\Lambda^*(u)\right)}e^{\beta^*X}\Lambda^*(u)\right\}\left\{dG\left(\int_0^u e^{\theta(s)X}d\Lambda_0(s)\right) - dG\left(\int_0^u e^{\beta^*X}d\Lambda^*(s)\right)\right\}.$$

Suppose that the limit is zero for all t. Since X is binary, $G(\int_0^t e^{\theta(s)}d\Lambda_0(s)) = G(\int_0^t e^{\beta^*}d\Lambda^*(s))$ for all t. Thus $e^{\theta(t)}\lambda_0(t) = e^{\beta^*}\lambda^*(t)$. Let

$$\phi(t;\Lambda,\beta) = E\left[Y(t)e^{\beta X}\left\{G'\left(\int_0^\tau Y(s)e^{\beta X}d\Lambda(s)\right) - \int_t^\tau \frac{G''}{G'}\left(\int_0^s Y(u)e^{\beta X}d\Lambda(u)\right)dN(s)\right\}\right].$$

By (A3) of Zeng and Lin (2006) and (S.8) in the proof of Lemma S.3, $\lambda_0(t)\phi(t;\Lambda_0,\theta(t)) = \lambda^*(t)\phi(t;\Lambda^*,\beta^*)$. This result, combined with the fact that $e^{\theta(t)}\lambda_0(t) = e^{\beta^*}\lambda^*(t)$, implies that

$$\frac{E\{f(s,\Lambda_0)|X=0\}}{E\{f(s,\Lambda^*)|X=0\}} = \frac{\lambda^*(t)}{\lambda_0(t)}, \tag{S.11}$$

where

$$f(t,\Lambda) = Y(t)\left[G'\left(\int_0^\tau Y(s)d\Lambda(s)\right) - \int_t^\tau \frac{G''(\Lambda(s))}{G'(\Lambda(s))}Y(s)dN(s)\right].$$

Since $\Lambda_0(t)$ is a solution to the score equation for Λ conditional on $X=0$, we obtain $\Lambda_0(t) = \int_0^t E\{dN(s)|X=0\}/E\{f(s,\Lambda_0)|X=0\}$. It then follows from (S.11) that $\Lambda^*(t) =$

$\int_0^t E\{dN(s)|X = 0\}/E\{f(s,\Lambda^*)|X = 0\}$, so $\Lambda^*(t)$ is also a solution to the score equation for Λ conditional on $X = 0$. Thus, $G(\Lambda_0(t))$ and $G(\Lambda^*(t))$ are the solutions to the same score equation. Since the score equation has an unique solution given by $\int_0^t E\{dN(s)|X = 0\}/E\{Y(s)|X = 0\}$, we obtain $G(\Lambda_0(t)) = G(\Lambda^*(t))$. This result, together with the fact that $e^{\theta(t)}\lambda_0(t) = e^{\beta^*}\lambda^*(t)$, implies $\theta(t) = \beta^*$, which is a contradiction.

S.3.5. Transformation function. The limit of $n^{-1/2}W_{tr}(x,t)$ under the alternative is

$$E\left[\int_0^t I(\Lambda^*(u)e^{\beta^{*T}X} \le x)Y(u)\left\{dG_0(\Lambda_0(u)e^{\beta_0^{\mathrm{T}}X}) - dG(\Lambda^*(u)e^{\beta^{*T}X})\right\}\right]. \tag{S.12}$$

Suppose that (S.12) is zero for all $x > 0$ and $t > 0$. By taking the derivative of (S.12) with respect to t, we have

$$E\left[I(\Lambda^*(t)e^{\beta^{*T}X} \le x)Y(t)\left\{G_0'(\Lambda_0(t)e^{\beta_0^{\mathrm{T}}X})\lambda_0(t)e^{\beta_0^{\mathrm{T}}X} - G'(\Lambda^*(t)e^{\beta^{*T}X})\lambda^*(t)e^{\beta^{*T}X}\right\}\right] = 0.$$

For every $y > 0$, let $x = y\Lambda^*(t)$. Then

$$E\left[I(e^{\beta^{*T}X} \le y)Y(t)\left\{G_0'(\Lambda_0(t)e^{\beta_0^{\mathrm{T}}X})\lambda_0(t)e^{\beta_0^{\mathrm{T}}X} - G'(\Lambda^*(t)e^{\beta^{*T}X})\lambda^*(t)e^{\beta^{*T}X}\right\}\right] = 0. \tag{S.13}$$

Letting $t \to 0$ and noticing that $G'(0) = G_0'(0)$ and $\lambda^*(0) = \lambda_0(0)$, we have $E\left[e^{\beta_0^{\mathrm{T}}X}\middle|e^{\beta^{*T}X} = y\right] = y$. Thus, $\beta_0 = \beta^*$. It then follows from (S.13) that for all $t > 0$ and $y > 0$,

$$G_0'(\Lambda_0(t)y)\lambda_0(t) = G'(\Lambda^*(t)y)\lambda^*(t),$$

which entails that $\Lambda^*(t) = c\Lambda_0(t)$ for some constant c. Since $\lambda^*(0) = \lambda_0(0)$, the constant c must be 1. We then conclude that $G_0(x) = G(x)$, which is a contradiction.

S.4. Transformation mean models

By extending the arguments of Lin et al. (2000) and Zeng and Lin (2006), we can prove that the maximum pseduo-likelihood estimators $\widehat{\beta}$ and $\widehat{\mu}(\cdot)$ converge to the true parameter values β_0 and $\mu_0(\cdot)$ under model (1.3). However, the variance estimators for $\widehat{\beta}$ and $\widehat{\mu}(\cdot)$ described in Zeng and Lin (2006) are not consistent when the recurrent event times within the same subject are dependent. To obtain consistent variance estimators, we establish that $\sqrt{n}((\widehat{\beta} - \beta_0)^{\mathrm{T}}\widetilde{h}_1 +$

$\int_0^\tau \widetilde{h}_2(u)d(\widehat{\mu} - \mu_0)(u))$ is asymptotically equivalent to $\sqrt{n}(\mathcal{P}_n - P_0)(S_{\beta_0}, S_{\mu_0})I^{-1}_{\beta_0,\mu_0}(\widetilde{h}_1, \widetilde{h}_2(\cdot))$ for all $(\widetilde{h}_1, \widetilde{h}_2(\cdot)) \in \mathcal{R}^p \times \mathcal{F}$, where $\mathcal{F}$ is defined in Section S.1, S_{β_0} and S_{μ_0} are the pseudo-likelihood score operators, and I_{β_0,μ_0} is the pseudo-likelihood information operator. Indeed, a similar representation is used in Section S.1. for $\widehat{\beta}$ and $\widehat{\Lambda}$. In light of this representation, together with the fact that $E\{dM(t; \beta_0, \mu_0))|X(\cdot)\} = 0$, we can show that the weak convergence of $W_n(x, t)$ described in Section 2 of the main paper (with Λ replaced by μ) holds under model (1.3). Consequently, all the theoretical results we have established for model (1.2) are applicable to model (1.3).

S.5. PBC sequential data

We consider the primary biliary cirrhosis (PBC) sequential database, which is a follow-up of the PBC study reported in Fleming and Harrington (1991). The PBC study was originally designed to evaluate the effect of the drug D-penicillamine on the survival time of PBC patients. The drug turned out to be ineffective, and the data were used to build a proportional hazards model for the natural history of PBC with covariates age, edema, log(bilirubin), log(albumin), and log(protime) (Fleming and Harrington, 1991). The PBC sequential database contains the follow-up values on the above five covariates for the 312 patients involved in the randomized clinical trial. A total of 140 patients had died by the end of follow-up.

We start with the proportional hazards model with five time-dependent covariates: age, edema, log(bilirubin), log(albumin), and log(pro-time). The supremum tests $\sup_x |W_c(x, \infty)|$ for checking the functional forms of the five covariates have p-values of .360, .447, .037, .011, and .419, respectively, indicating that the functional forms of bilirubin and albumin may be inappropriate. Figure S3 plots the cumulative sums of residuals $W_c(\cdot, \infty)$ for log(bilirubin). The observed pattern resembles the solid curve of Figure 1b, suggesting the addition of square term of log(bilirubin). The plot of the cumulative sums of residuals $W_c(\cdot, \infty)$ for log(albumin) (omitted here) suggests adding square and cubic terms of log(albumin). After adding these terms, the p-values of the supremum tests for the functional forms of log(bilirubin) and log(albumin) increase to .380 and .075, respectively. The p-values of the supremum tests for the functional forms of age, edema, and log(protime) become .326, .468, and .581, respectively. The supremum test $\sup_x |W_l(x, \infty)|$ for checking the link function has a p-value of .728. The supremum tests $\sup_t |W_p(t)|$ for checking the proportionality of age, edema, log(bilirubin), log(albumin), log(protime), square of log(bilirubin), and square and cubic of log(albumin) have p-values of .759, .048, .657, .553, .450, .220, .327, .200, indicating that the proportional hazards assumption for edema may be problematic. Figure S4 displays the score process $W_p(\cdot)$ for edema. The observed curve is concave and above zero, and the estimated regression parameter for edema is positive. According to Figure 2, this pattern indicates that the hazards ratio for two edema values decreases over time. The supremum tests $\sup_x |W_{tr}(x, \infty)|$ and $\sup_{x,t} |W_{tr}(x, t)|$ for checking the transformation function have p-values of .037 and .081. To correct the non-proportionality of edema, we add the interaction of edema and $\log t$ to the model. For the new model, the supremum tests for the functional forms of age, edema, log(bilirubin), log(albumin), log(protime), and edema$\times \log t$ have p-values of .372, .501, .319, .072, .555, and .842, respectively; the supremum test for the link function has a p-value of .736; the supremum tests for the proportionality of age, edema, log(bilirubin), log(albumin), log(protime), square of log(bilirubin), and square and cubic of log(albumin) have p-values of .778, .448, .270, .668, .215, .077, .685, .494, .723, respectively; the supremum tests $\sup_x |W_{tr}(x, \infty)|$ and $\sup_x |W_{tr}(x, t)|$ have p-values of .083 and .134, respectively. The above p-values show that the new model is satisfactory. Table S6 shows the estimation results for the final model. The square term of log(bilirubin) and the cubic term of log(albumin) are very significant. The interaction between edema and $\log t$ is significant and has a negative coefficient, showing that the hazards ratio associated with two edema values decreases to 1 over time.

Additional References

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* New York: Springer.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society*, Series B, **62**, 711–730.

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

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93,** 627–640.

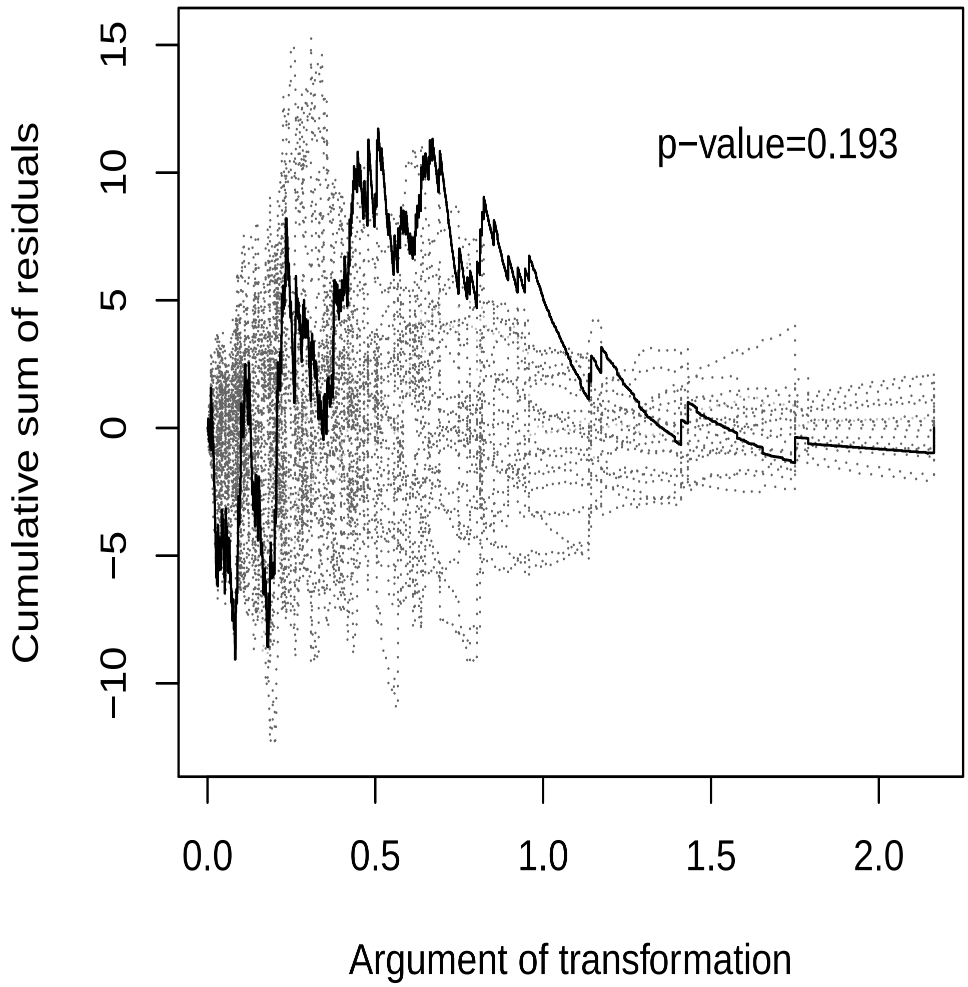

Figure S1: Plot of the cumulative sum of residuals $W_{tr}(\cdot, \infty)$ in the colon cancer data: the observed pattern is shown by the solid curve while 20 simulated realizations from the null distribution are shown in dotted curves.

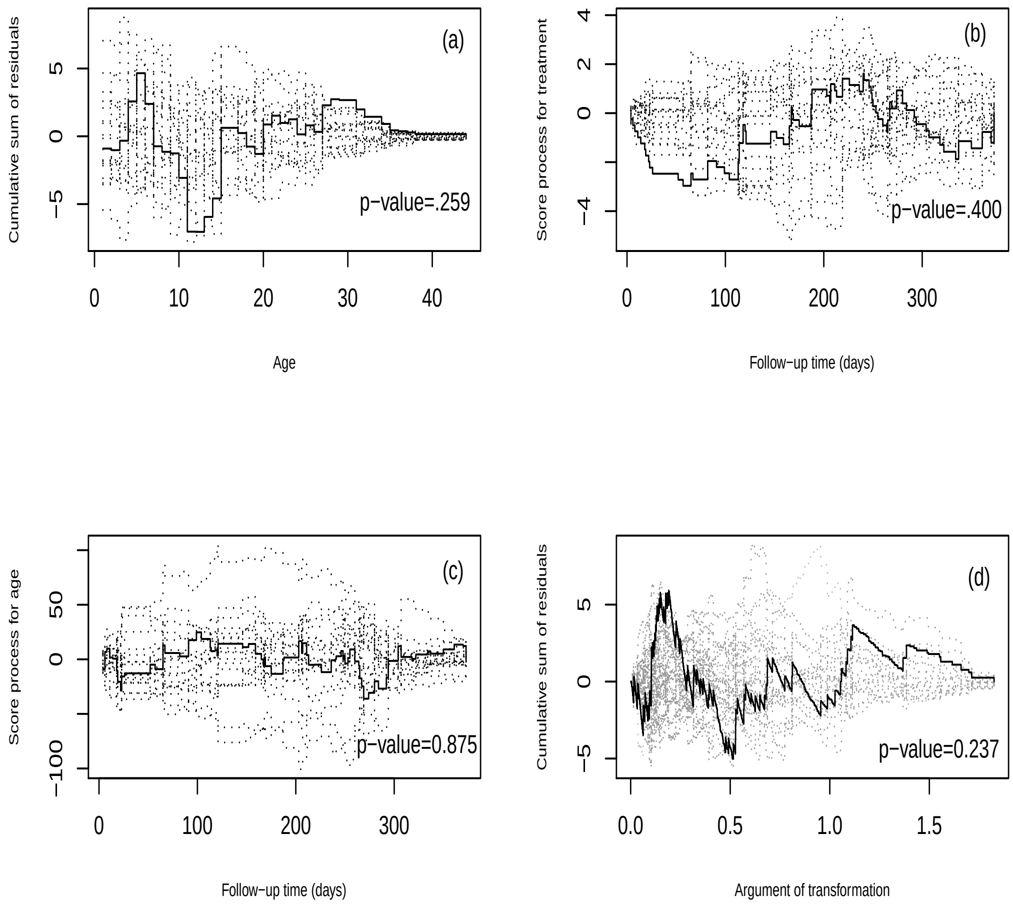

Figure S2: Residual plots under the proportional means model for the CGD data: (a) cumulative sum of residuals $W_c(\cdot,\infty)$ for the functional form of age; (b) score process $W_p(\cdot)$ for treatment; (c) score process $W_p(\cdot)$ for age; and (d) cumulative sum of residuals $W_{tr}(\cdot,\infty)$ for checking the transformation function. In each plot, the observed pattern is shown by the solid curve while 20 simulated realizations from the null distribution are shown in dotted curves.

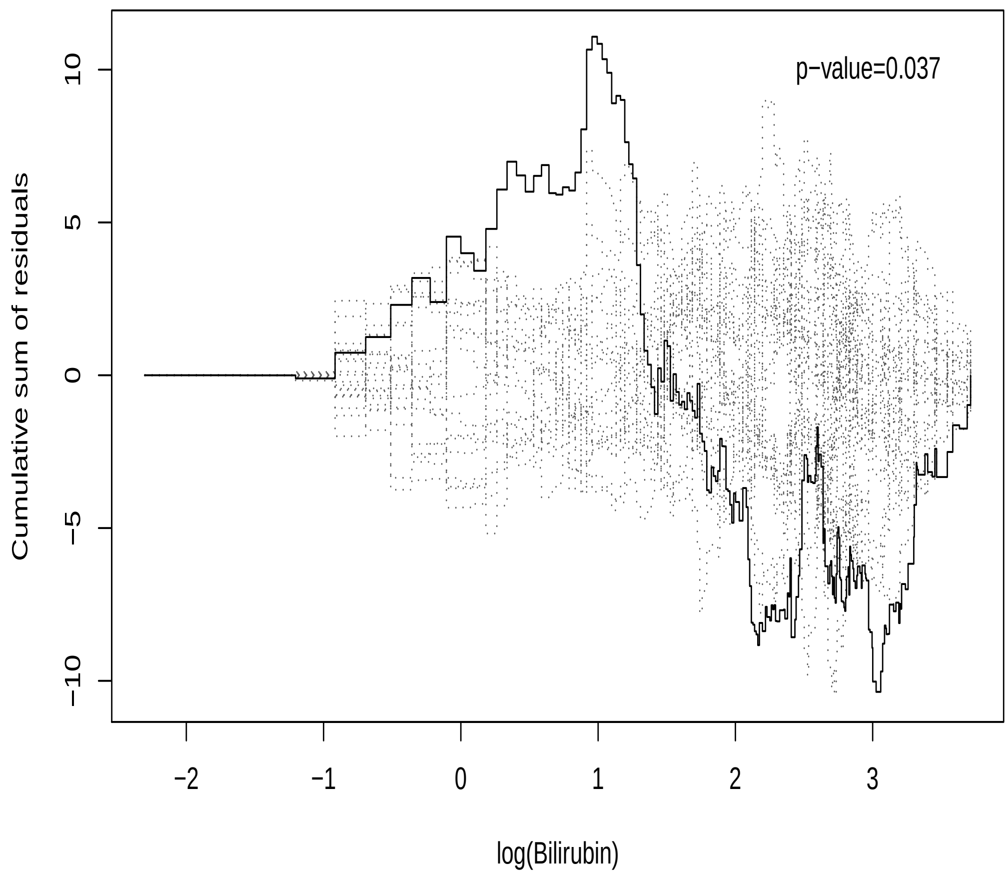

Figure S3: Plot of the cumulative sum of residuals $W_c(\cdot, \infty)$ for the functional form of time-dependent log(bilirubin) in the PBC sequential data: the observed pattern is shown by the solid curve while 20 simulated realizations from the null distribution are shown in dotted curves.

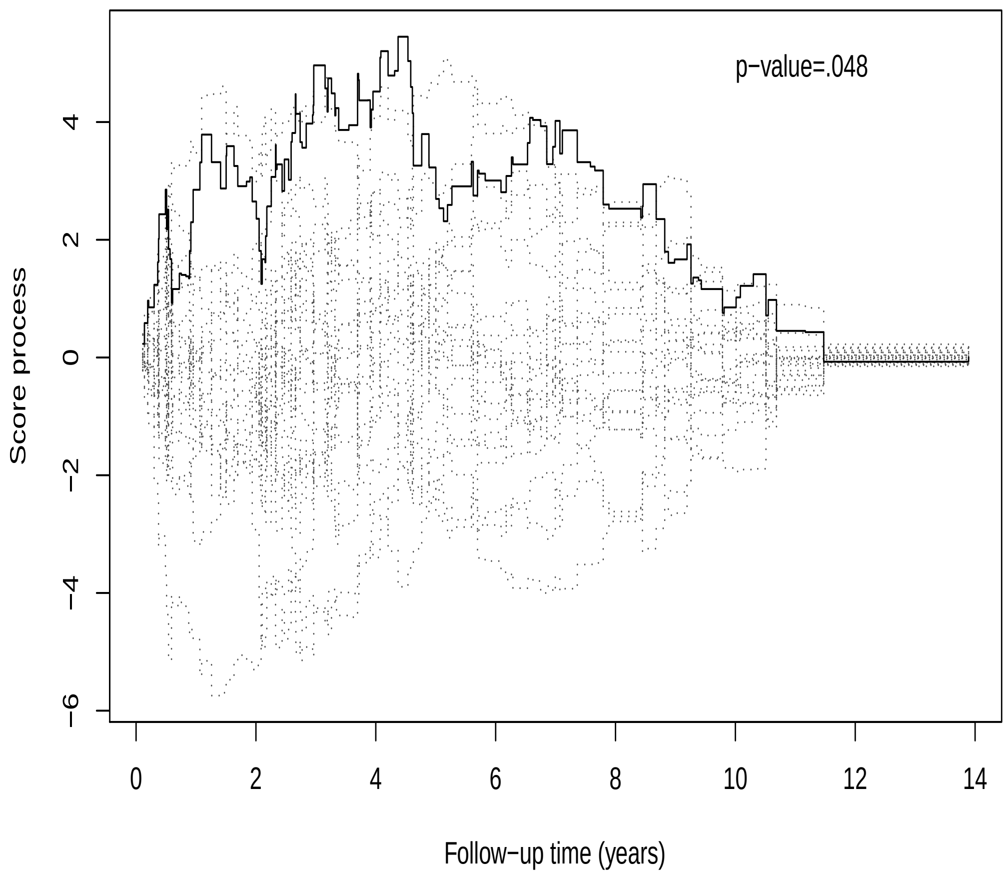

Figure S4: Plot of the score process $W_p(\cdot)$ for time-dependent edema in the PBC sequential data: the observed pattern is shown by the solid curve while 20 simulated realizations from the null distribution are shown in dotted curves.

Table S1: Type I error of the supremum tests for survival data and recurrent events with time-independent covariates at the 5% significance level with n=200 (100)

	Survival Data					
	Box-Cox			logarithmic		
	$\rho = .5$	1	2	$r = .5$	1	2
censoring	28%	20%	9%	30%	37%	48%
$\sup_{x,t} \lvert W_c^{(2)}(x,t)\rvert$	.052 (.056)	.042 (.050)	.054 (.046)	.055 (.056)	.055 (.071)	.055 (.085)
$\sup_{x} \lvert W_c^{(2)}(x,\infty)\rvert$	.051 (.056)	.046 (.046)	.055 (.056)	.054 (.065)	.053 (.066)	.053 (.078)
$\sup_{x,t} \lvert W_l(x,t)\rvert$	.042 (.066)	.042 (.049)	.039 (.041)	.045 (.062)	.054 (.072)	.056 (.084)
$\sup_{x} \lvert W_l(x,\infty)\rvert$	.042 (.059)	.038 (.057)	.040 (.051)	.040 (.056)	.053 (.076)	.059 (.089)
$\sup_{t} \lvert W_p^{(1)}(t)\rvert$	.052 (.063)	.053 (.068)	.054 (.068)	.054 (.056)	.049 (.051)	.047 (.053)
$\sup_{t} \lvert W_p^{(2)}(t)\rvert$	.044 (.049)	.049 (.052)	.058 (.049)	.044 (.048)	.049 (.052)	.049 (.050)
$\sup_{x,t} \lvert W_{tr}(x,t)\rvert$	.054 (.051)	.047 (.049)	.050 (.056)	.050 (.049)	.049 (.056)	.072 (.077)
$\sup_{x} \lvert W_{tr}(x,\infty)\rvert$	.045 (.048)	.042 (.052)	.051 (.055)	.047 (.044)	.045 (.056)	.067 (.078)
$\sup_{x,t} \lvert W_o(x,t)\rvert$	.056 (.059)	.047 (.054)	.054 (.047)	.054 (.059)	.055 (.070)	.064 (.079)
	Recurrent Event Data					
	Box-Cox			logarithmic		
	$\rho = .5$	1	2	$r = .5$	1	2
# events/subject	1.41	1.97	4.36	1.33	1.04	0.77
$\sup_{x,t} \lvert W_c^{(2)}(x,t)\rvert$	.048 (.044)	.043 (.042)	.048 (.040)	.048 (.042)	.043 (.036)	.048 (.039)
$\sup_{x} \lvert W_c^{(2)}(x,\infty)\rvert$	.052 (.036)	.038(.044)	.045 (.041)	.048 (.040)	.043 (.040)	.049 (.037)
$\sup_{x,t} \lvert W_l(x,t)\rvert$	.045 (.044)	.037 (.035)	.043(.038)	.055 (.047)	.044 (.042)	.051 (.048)
$\sup_{x} \lvert W_l(x,\infty)\rvert$	.048 (.046)	.036 (.040)	.046 (.041)	.052 (.040)	.046 (.042)	.052 (.047)
$\sup_{t} \lvert W_p^{(1)}(t)\rvert$	.044 (.053)	.052 (.051)	.050 (.048)	.048 (.050)	.049 (.064)	.056 (.060)
$\sup_{t} \lvert W_p^{(2)}(t)\rvert$	.054 (.041)	.051 (.042)	.044 (.036)	.036 (.047)	.047 (.044)	.058 (.037)
$\sup_{x,t} \lvert W_{tr}(x,t)\rvert$	.057 (.051)	.055 (.051)	.045 (.042)	.053 (.053)	.053 (.060)	.062 (.070)
$\sup_{x} \lvert W_{tr}(x,\infty)\rvert$	.053 (.045)	.052 (.051)	.040 (.043)	.051 (.047)	.054 (.058)	.057 (.059)
$\sup_{x,t} \lvert W_o(x,t)\rvert$	.046 (.041)	.043 (.041)	.047 (.040)	.047 (.042)	.043 (.035)	.047 (.037)

Note: For survival data, Box-Cox with $\rho = 1$ is the proportional hazards model and logarithmic with $r = 1$ is the proportional odds model. Here and in the sequel, the results are based on 2000 replicates.

Table S2: Type I error of the supremum tests for survival data with time-dependent covariates at the 5% significance level with n=200 (100)

	Box-Cox			logarithmic		
	$\rho = .5$	1	2	$r = .5$	1	2
censoring	23%	15%	7%	25%	32%	43%
$\sup_{x,t} \lvert W_c^{(2)}(x,t)\rvert$	.060 (.066)	.065 (.066)	.064 (.070)	.068 (.069)	.064 (.071)	.063 (.071)
$\sup_{x} \lvert W_c^{(2)}(x,\infty)\rvert$	.056 (.055)	.062 (.061)	.064 (.066)	.054 (.067)	.060 (.069)	.068 (.070)
$\sup_{x,t} \lvert W_l(x,t)\rvert$	.057 (.055)	.050 (.064)	.052 (.058)	.048 (.059)	.058 (.060)	.062 (.074)
$\sup_{x} \lvert W_l(x,\infty)\rvert$	.060 (.051)	.052 (.060)	.066 (.070)	.055 (.062)	.064 (.062)	.069 (.072)
$\sup_{t} \lvert W_p^{(1)}(t)\rvert$	.054 (.060)	.054 (.059)	.059 (.067)	.051 (.055)	.056 (.051)	.050 (.053)
$\sup_{t} \lvert W_p^{(2)}(t)\rvert$	.052 (.047)	.064 (.053)	.051 (.054)	.051 (.056)	.046 (.050)	.057 (.057)
$\sup_{x,t} \lvert W_{tr}(x,t)\rvert$	.050 (.055)	.058 (.063)	.069 (.070)	.051 (.052)	.061 (.061)	.074 (.086)
$\sup_{x} \lvert W_{tr}(x,\infty)\rvert$	.048 (.058)	.053 (.062)	.061 (.066)	.049 (.057)	.055 (.061)	.066 (.082)
$\sup_{x,t} \lvert W_o(x,t)\rvert$	.060 (.069)	.064 (.067)	.064 (.071)	.066 (.070)	.062 (.070)	.066 (.070)

Note: Box-Cox with $\rho = 1$ is the proportional hazards model and logarithmic with $r = 1$ is the proportional odds model.

Table S3: Power of the supremum and Wald tests at the 5% significance level against the omission of a squared term

	Survival Data					
	Box-Cox(n=100)			logarithmic(n=200)		
	$\rho = .5$	1	2	$r = .5$	1	2
τ	3.7	2.2	1.2	1.0	1.4	3.0
censoring	53%	58%	65%	77%	74%	68%
$\sup_{x,t} \lvert W_c(x,t)\rvert$	.651	.692	.695	.717	.687	.598
$\sup_{x} \lvert W_c(x,\infty)\rvert$	.784	.780	.742	.723	.662	.557
Wald	.901	.903	.894	.887	.897	.896

	Recurrent Events Data					
	Box-Cox(n=100)			logarithmic(n=200)		
	$\rho = .5$	1	2	$r = .5$	1	2
τ	3.7	2.2	1.2	1.0	1.4	3.0
# events/subject	0.94	0.80	0.63	0.31	0.36	0.45
$\sup_{x,t} \lvert W_c(x,t)\rvert$	.713	.727	.689	.663	.663	.542
$\sup_{x} \lvert W_c(x,\infty)\rvert$	.723	.792	.792	.691	.627	.481
Wald	.962	.963	.934	.922	.925	.935

Note: $\sup_{x,t} |W_c(x,t)|$ is the same as $\sup_{x,t} |W_o(x,t)|$ when there is only one covariate.

Table S4: Power of the supremum and Wald tests at the 5% significance level against a logistic functional form of the covariate

	Box-Cox(n=100)			logarithmic(n=200)		
	$\rho=.5$	1	2	$r=.5$	1	2
τ	3	3	1	1.5	2	2
censoring	17%	13%	24%	26%	27%	36%
$\sup_{x,t}\lvert W_c(x,t)\rvert$	.764	.885	.941	.952	.841	.490
$\sup_{x}\lvert W_c(x,\infty)\rvert$	.841	.949	.984	.962	.825	.421
Wald	.192	.261	.120	.155	.174	.141

Table S5: Power of the supremum and Wald tests against nonproportionality

	$\Lambda(t\vert X)=G(\int_0^t \exp(.2X+.2X*\log(s))ds)$					
	Box-Cox ($n=100$)			logarithmic ($n=100$)		
	$\rho=.5$	1	2	$r=.5$	1	2
censoring	27%	23%	18%	28%	34%	44%
$\sup_t \lvert W_p^{(1)}(t)\rvert$	.742	.784	.792	.688	.605	.604
Wald	.889	.908	.917	.868	.818	.728
	$\Lambda(t\vert X)=G(\int_0^t \exp(-.1X-.5X*\sin(2*s))ds)$					
	Box-Cox ($n=100$)			logarithmic ($n=200$)		
	$\rho=.5$	1	2	$r=.5$	1	2
censoring	45%	48%	53%	44%	40%	35%
$\sup_t \lvert W_p^{(1)}(t)\rvert$	.900	.944	.949	.993	.959	.752
Wald	.369	.460	.512	.499	.345	.192

Table S6: Analysis of the PBC sequential data under the proportional hazards model

Parameter	Est	SE	Est/SE	p-value
Age	0.053	0.010	5.573	<.001
Edema	1.382	0.376	3.678	<.001
log(Bilirubin)	0.238	0.239	0.996	.319
log(Albumin)	9.817	6.006	1.635	.102
log(Protime)	2.448	0.673	3.638	<.001
$\log^2$(Bilirubin)	0.277	0.076	3.634	<.001
$\log^2$(Albumin)	−16.360	5.890	−2.778	.005
$\log^3$(Albumin)	5.992	1.870	3.205	.001
Edema$*\log t$	−0.559	0.247	−2.260	.024