Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring

Xingqiu Zhao[1,*], Li Liu[2], Yanyan Liu[2] and Wei Xu[3]

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

[2]School of Mathematics and Statistics, Wuhan University, Wuhan, China

[3]Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Supplementary Information

For sake of convenience, let $\sum_{i<l}$ stand for $\sum_{i=1}^{n}\sum_{l=i+1}^{n}$ and $\sum_{i<l_1<l_2}$ stand for $\sum_{i=1}^{n}\sum_{l_1=i+1}^{n}\sum_{l_2=l_1+1}^{n}$.

Proof of Theorem 1

Since $-\log(1+\rho_{ij,lj}(T_{ijk},T_{ljs})'\beta)$ is the log-likelihood of T_{ijk} and T_{ljs} conditional on

$$\{(\xi_{i0},C_{ij},Z_{ij},m_{ij},\mathcal{X}_{ij}(C_{ij})),(C_{lj},Z_{lj},m_{lj},\mathcal{X}_{lj}(C_{lj}))\}$$

and the order statistics of T_{ijk} and T_{ljs}, it follows that $-\log(1+\rho_{ij,lj}(T_{ijk},T_{ljs})'\beta)$ achieves its maximum at the true parameter value β_0. By the conditional Kullback-Leibler information inequality (Andersen, 1970), the maximum pairwise pseudolikelihood estimator $\hat{\beta}$ is consistent.

Next, we derive the asymptotic normality of $\hat{\beta}$. Applying Taylor expansion to $l(\beta)$ gives

$$\sqrt{n}(\hat{\beta}-\beta_0)=(-\partial l(\beta_0)/\partial\beta_0)^{-1}\cdot\sqrt{n}l(\beta_0)+o_p(1), \tag{1}$$

*Corresponding author: e-mail: xingqiu.zhao@polyu.edu.hk, Phone: +852-2766-6921, Fax: +852-2362-9045

where

$$\begin{aligned}
\frac{\partial l(\beta)}{\partial \beta} &= \frac{1}{\binom{n}{2}} \sum_{i<l} \frac{\partial H(D_i, D_l; \beta)}{\partial \beta} \\
&= \frac{1}{\binom{n}{2}} \sum_{i<l} \sum_{j=1}^{J} \iint_0^{C_{ij} \wedge C_{lj}} \frac{\exp\{\rho_{ij,lj}(t,u)'\beta\}}{(1+\exp\{\rho_{ij,lj}(t,u)'\beta\})^2} \rho_{ij,lj}^2(t,u) dN_{ij}(t) dN_{lj}(u)
\end{aligned}$$

is also a U-statistic. It is easy to show that $\sqrt{n}l(\beta_0)$ converges weakly to a normal distribution with mean 0 and variance covariance $V_1(\beta_0)$, and $\partial l(\beta_0)/\partial \beta_0$ converges almost surely to $V_2(\beta_0)$. Hence, Theorem 1 follows from (1).

Proof of Lemma 1

For any $u \in [0, \tau]$ and $d = 0, 1$, define

$$\tilde{Q}_{0j}^d(u;\beta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m_{ij}} I(T_{ijk} \le u) \exp(-X_{ij}(T_{ijk})'\beta) X_{ij}(T_{ijk})^d, \quad \text{and}$$

$$\tilde{R}_{0j}^d(u;\beta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m_{ij}} I(T_{ijk} \le u \le C_{ij}) \exp\{-X_{ij}(T_{ijk})'\beta\} X_{ij}(T_{ijk})^d.$$

It is easy to see that

$$\int_t^\tau \frac{dQ_{0j}^0(u)}{R_{0j}^0(u)} = \int_t^\tau \frac{d\Lambda_{0j}(u)}{\Lambda_{0j}(u)} = -\ln\left(\frac{\Lambda_{0j}(t)}{\Lambda_{0j}(\tau)}\right) = -\ln F_{0j}(t).$$

By using similar arguments used in the proof of Lemma 1 of Huang, Qin and Wang (2010), we have for each $t \in [c^*, \tau]$,

$$\hat{F}_{0j}(t) = \exp\left(-\int_t^\tau \frac{d\tilde{Q}_{0j}^0(u;\hat{\beta})}{\tilde{R}_{0j}^0(u;\hat{\beta})}\right) + o_p(n^{-1/2}),$$

and

$$\begin{aligned}
&\int_t^\tau \frac{d\tilde{Q}_{0j}^0(u;\hat{\beta})}{\tilde{R}_{0j}^0(u;\hat{\beta})} - \int_t^\tau \frac{dQ_{0j}^0(u)}{R_{0j}^0(u)} \\
=\ & \frac{1}{\binom{n}{2}} \sum_{i<l} \zeta_{0j}(D_i, D_l; t, \beta_0) + \frac{1}{n} \sum_{i=1}^{n} \psi_{0j}(D_i; t, \beta_0) + o_p(n^{-1/2}).
\end{aligned}$$

Note that $\frac{1}{\binom{n}{2}}\sum_{i<l}\kappa_{0j}(D_i, D_l; t, \beta)$ is a U-statistic. Similar to the proof of Theorem 1, for each fixed $t \in [c^*, \tau]$, $\sqrt{n}\left(\int_t^\tau \frac{d\tilde{Q}_{0j}^0(u;\hat{\beta})}{\tilde{R}_{0j}^0(u;\hat{\beta})} - \int_t^\tau \frac{dQ_{0j}^0(u)}{R_{0j}^0(u)}, j = 1, \ldots, J\right)'$ converges weakly to a normal distribution with mean 0 and variance $4E[\kappa_0(D_1, D_2; t, \beta_0)\kappa_0(D_1, D_3; t, \beta_0)']$, where

$$\kappa_0(D_i, D_l; t, \beta_0) = (\kappa_{0j}(D_i, D_l; t, \beta_0), j = 1, \ldots, J)'.$$

Applying the delta method, we have

$$\hat{F}_{0j}(t) - F_{0j}(t) = -\frac{1}{\binom{n}{2}}\sum_{i<l}\kappa_{0j}(D_i, D_l; t, \beta_0)F_{0j}(t) + o_p(n^{-1/2}),$$

and for fixed $t \in (\tau_0, \tau]$, $\sqrt{n}(\hat{F}_{0j}(t) - F_{0j}(t), j = 1, \ldots, J)'$ converges weakly to a normal distribution with mean 0 and variance $\Sigma_\kappa(t)$.

Proof of Theorem 2

Following the strong consistency of $\hat{\beta}$ and $\hat{F}_{0j}(\cdot)$, we have

$$\begin{aligned} M_{ij}(\hat{\beta}, \hat{F}_{0j}) &= \left[\frac{m_{ij}}{\int_0^{C_{ij}} \exp\{X_{ij}(u)'\hat{\beta}\}d\hat{F}_{0j}(u)} - \frac{m_{ij}}{\int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta_0\}d\hat{F}_{0j}(u)}\right] \\ &\quad + \frac{m_{ij}}{\int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta_0\}dF_{0j}(u)} \\ &= \frac{-m_{ij}}{\left\{\int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta_0\}dF_{0j}(u)\right\}^2}\left[\int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta_0\}X_{ij}(u)'dF_{0j}(u)(\hat{\beta} - \beta_0)\right. \\ &\quad \left. + \int_0^{C_{ij}} \exp\{X_{ij}(u)'\beta_0\}d\{\hat{F}_{0j}(u) - F_{0j}(u)\}\right] + M_{ij}(\beta_0, F_{0j}) + o_p(n^{-1/2}). \qquad (2) \end{aligned}$$

We first show the asymptotic properties of $\hat{\eta}$. From the proofs of Theorem 1 and Lemma 1, it is esay to see that

$$\frac{1}{n}\sum_{i=1}^n Z_i^* M_i(\hat{\eta}; \hat{\beta}, \hat{F}) = \frac{1}{\binom{n}{2}}\sum_{i<l}\iota(D_i, D_l) + o_p(n^{-1/2}),$$

and $\frac{1}{\binom{n}{2}}\sum_{i<l}\iota(D_i, D_l)$ is a U-statistics. Similar to the arguments as that of the proof in Theorem 1, we have $\sqrt{n}(\hat{\eta} - \eta)$ converges to the multivariate normal distribution with mean

0 and variance-covariance matrix Σ_η. Hence, it follows that $\sqrt{n}(\hat{\gamma} - \gamma_0)$ converges weakly to a normal distribution with mean 0 and variance Σ_γ.

To derive the asymptotic normality of $\sqrt{n}(\hat{\Lambda}_{0j}(t) - \Lambda_{0j}(t), j = 1, \ldots, J)'$, we write

$$\sqrt{n}(\hat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)) \quad = \quad \sqrt{n} F_{0j}(t) \exp(\eta_j) \frac{1}{\binom{n}{2}} \sum_{i<l} \{ f_{0j}(D_i, D_l) + \kappa_{0j}(D_i, D_l; t, \beta_0) \} + o_p(1).$$

Hence, by the central limit theorem for U-statistics, $\sqrt{n}(\hat{\Lambda}_{0j}(t) - \Lambda_{0j}(t), j = 1, \ldots, J)'$ converges weakly to the multivariate normal distribution with mean 0 and covariance $\Sigma_\Lambda(t)$.

Proof of Theorem 3

Now we turn to the asymptotic properties of $\tilde{\eta}$. By using (2), we can get

$$\frac{1}{n} \sum_{i=1}^{n} Z_i^* \tilde{M}_i(\tilde{\eta}; \hat{\beta}, \hat{F}) = \frac{1}{\binom{n}{2}} \sum_{i<l} \tilde{\iota}(D_i, D_l) + o(n^{-1/2}).$$

It can be verified that $\frac{1}{\binom{n}{2}} \sum_{i \neq l} \tilde{\iota}(D_i, D_l)$ is a U-statistics. Thus, by the same arguments as the proof of Theorem 2, we can obtain the results of Theorem 3.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B* **32**, 283-301.

Huang, C. Y., Qin, J. and Wang, M. C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics* **66**, 39-49.

Table 1: Simulation results based on Schaubel and Cai's method and the proposed method with $n = 200$, where censoring time $C_{ij} = \min(10, C_{ij}^1)$ with $C_{ij}^1 \sim E(1/10)$ if $Z_{ij} = 1$, $C_{ij}^1 | (\xi_{i0}, \mathcal{X}_{ij}, Z_{ij}) \sim E(1/(6\xi_{i0} + 4))$ if $Z_{ij} = 0$

			$\bar{\beta}$		$\bar{\gamma}$		$\hat{\beta}$		$\hat{\gamma}$	
β_0	γ_0	σ	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
0	0.5	1	0.004	0.128	−0.186	0.136	−0.002	0.107	−0.029	0.147
		0.7	0.003	0.103	−0.108	0.103	0.007	0.107	−0.023	0.124
		0.5	−0.001	0.093	−0.057	0.084	0.000	0.106	−0.016	0.106
		0.3	−0.002	0.075	−0.021	0.068	−0.002	0.113	0.001	0.105
0.2	0.5	1	−0.003	0.130	−0.200	0.135	−0.010	0.114	−0.035	0.152
		0.7	0.001	0.103	−0.115	0.103	0.007	0.113	−0.024	0.125
		0.5	−0.001	0.093	−0.061	0.082	0.002	0.109	−0.018	0.113
		0.3	−0.002	0.072	−0.022	0.066	−0.001	0.117	0.001	0.106
0.2	−0.3	1	0.002	0.137	−0.190	0.143	−0.004	0.136	−0.026	0.164
		0.7	0.001	0.109	−0.115	0.110	0.008	0.133	−0.022	0.138
		0.5	0.003	0.101	−0.061	0.092	0.007	0.138	−0.014	0.130
		0.3	0.001	0.081	−0.025	0.076	0.003	0.141	−0.004	0.121

Note: $\bar{\beta}$ and $\bar{\gamma}$ represent the estimates of regression parameters β and γ in Schaubel and Cai (2005). SSE represents the sample standard error of estimates.

Table 2: Simulation results based on Schaubel and Cai's method and the proposed method with $n = 200$, where censoring time $C_{ij} = \min(10, C^1_{ij})$ with $C^1_{ij}|(\xi_{i0}, \mathcal{X}_{ij}, Z_{ij}) \sim E(1/(6\xi_{i0}+4))$ if $Z_{ij} = 1$ or $X_{ij} \leq 0.5$, $C^1_{ij} \sim E(1/10)$ otherwise

			$\bar{\beta}$		$\bar{\gamma}$		$\hat{\beta}$		$\hat{\gamma}$	
β_0	γ_0	σ	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
0	0.5	1	−0.067	0.123	0.093	0.132	−0.002	0.106	0.008	0.148
		0.7	−0.038	0.098	0.048	0.101	0.009	0.104	0.001	0.125
		0.5	−0.022	0.086	0.028	0.084	−0.002	0.103	0.001	0.109
		0.3	−0.011	0.073	0.011	0.070	−0.002	0.112	0.004	0.096
0.2	0.5	1	−0.078	0.125	0.099	0.130	−0.011	0.114	0.005	0.148
		0.7	−0.041	0.099	0.055	0.100	0.008	0.109	0.002	0.129
		0.5	−0.023	0.087	0.032	0.083	−0.002	0.108	−0.002	0.113
		0.3	−0.011	0.069	0.012	0.067	0.001	0.116	0.003	0.099
0.2	−0.3	1	−0.109	0.137	0.114	0.137	−0.001	0.140	0.018	0.161
		0.7	−0.063	0.109	0.057	0.106	0.011	0.136	0.007	0.138
		0.5	−0.040	0.100	0.044	0.107	0.004	0.134	0.000	0.130
		0.3	−0.013	0.080	0.010	0.079	0.004	0.139	0.002	0.122

Note: $\bar{\beta}$ and $\bar{\gamma}$ represent the estimates of regression parameters β and γ in Schaubel and Cai (2005). SSE represents the sample standard error of estimates.

Table 3: Estimation results of the regression coefficients for four types of transfusion reactions, i.e., fever, chill, rigor and all other symptoms. The corresponding $p-$values of the estimators are indicated in the parentheses

	$\hat{\beta}$		$\hat{\gamma}$		
Type	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\gamma}^{(1)}$	$\hat{\gamma}^{(2)}$	$\hat{\gamma}^{(3)}$
I	2.058 (< 0.01)	0.610 (< 0.01)	−0.203 (0.621)	−0.074 (0.845)	0.049 (0.803)
II	1.680 (< 0.01)	1.736 (< 0.01)	−0.386 (0.523)	0.269 (0.607)	0.370 (0.208)
III	1.828 (< 0.01)	0.569 (< 0.01)	−0.206 (0.689)	−0.348 (0.435)	0.352 (0.109)
IV	−0.608 (< 0.01)	2.405 (< 0.01)	−0.985 (0.032)	0.761 (0.044)	0.060 (0.750)

Type I = “Fever”, Type II = “Chill”, Type III = “Rigor”, and Type IV = “All other reactions”.

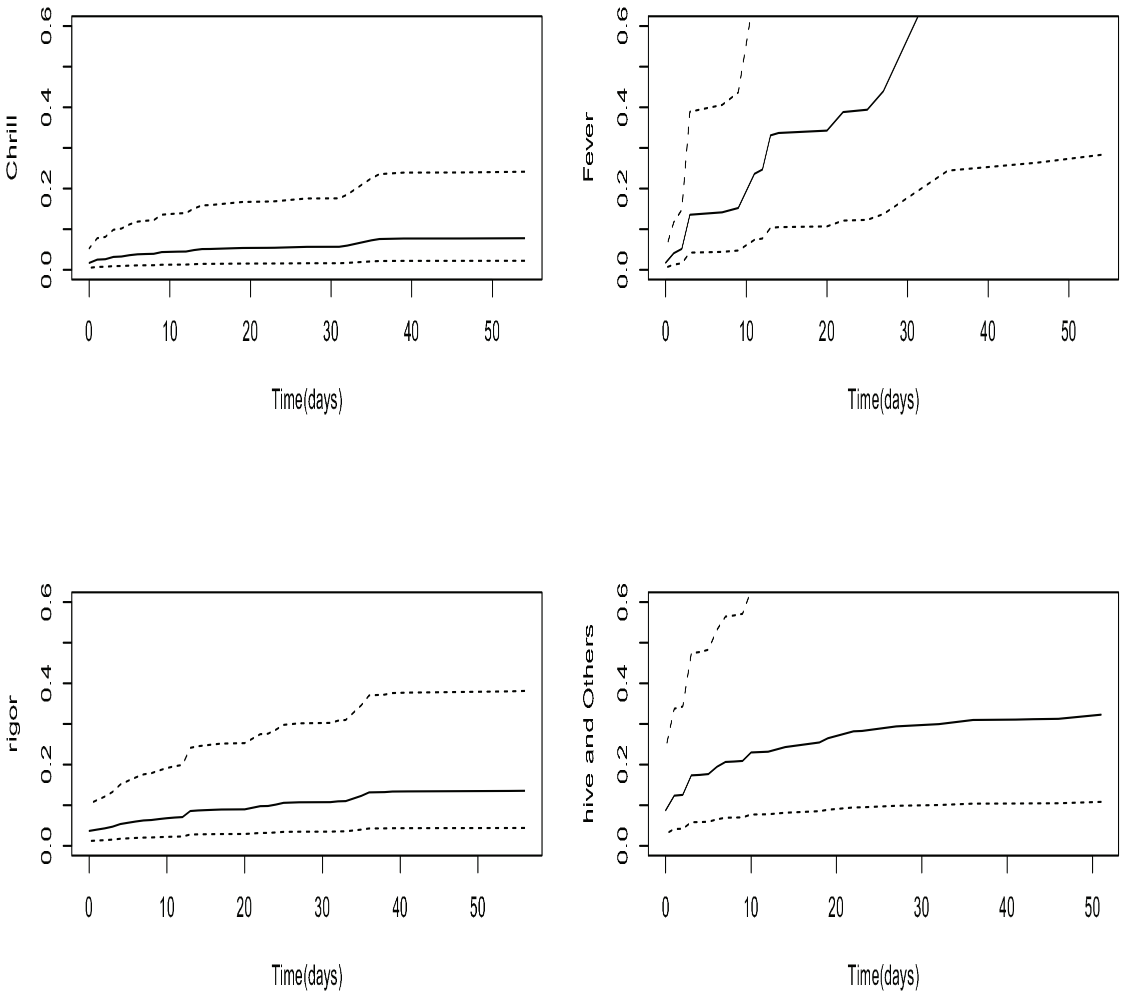

Figure 1 Plots of $\hat{\Lambda}_{0j}(t), j = 1, \ldots, 4$, the estimated mean numbers of recurrence of fever, chill, rigor and all other reactions in the FNHTR Data, with pointwise bootstrap 95% Confidence Intervals.