Stat Biosci (2017) 9:339–356
https://doi.org/10.1007/s12561-016-9161-0

Subdistribution Regression for Recurrent Events Under Competing Risks: with Application to Shunt Thrombosis Study in Dialysis Patients

Chia-Hui Huang[1] · Bowen Li[2] · Chyong-Mei Chen[3] · Weijing Wang[4] · Yi-Hau Chen[5]

Received: 14 October 2015 / Revised: 22 June 2016 / Accepted: 19 July 2016 /
Published online: 27 July 2016

Abstract This work is motivated by a nephrology study in Taiwan, where, after shunt implantation, dialysis patients may experience one of the two types, acute and non-acute, of shunt thrombosis, and each of them may alternatively recur in a patient. In this work, treating the two types of shunt thrombosis as competing risks, we assess covariate effects on the cumulative incidence probability function, or subdistribution, of gap times to the occurrences of acute shunt thrombosis. To accommodate potentially time-varying covariate effects, we extend a varying-coefficient subdistribution regression model to recurrent event analysis and propose associated estimation procedures. The inverse probability of censoring weighting technique is employed to ensure consistent estimation of the regression parameter. Asymptotic distributional theory is derived for the proposed estimator. Simulation results confirm that the proposed estimator performs well in finite samples. Application of the proposed analysis to the shunt thrombosis data reveals that dialysis patients with graft shunts and hypertension are associated with significantly increased incidence of acute shunt thrombosis.

Electronic supplementary material The online version of this article (doi:10.1007/s12561-016-9161-0) contains supplementary material, which is available to authorized users.

✉ Yi-Hau Chen
yhchen@stat.sinica.edu.tw

1 Department of Statistics, National Taipei University, Taipei, Taiwan

2 KKBOX Taiwan Co, Taipei, Taiwan

3 Institute of Public Health, School of Medicine, National Yang-Ming University, Taipei, Taiwan

4 Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

5 Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Keywords Cumulative incidence function · Gap times · Hemodialysis · Inverse probability weighting · Recurrent event · Time-varying coefficient

1 Introduction

Recurrent events are commonly encountered in biomedical studies. Examples include repeated infections after surgery and tumor recurrences. It is often of great interest to examine factors influencing times since enrollment to recurrent events and/or gap times between recurrent events. Regression analysis of recurrent events using either the calender time (time since enrollment), which is typically termed "total time" in the literature, or the gap time as the time index has drawn increasing attention; see for example Prentice et al. [10], Andersen and Gill [1], Wang et al. [17], Zeng and Lin [16] for analysis based on total time, and Huang and Chen [7], Schaubel and Cai [11], Lu [9], Sun et al. [15], and Huang and Liu [8] for that based on gap time. Cook and Lawless [4] provide an excellent review of recurrent event data analysis.

Our work is motivated by a nephrology study conducted by a large hospital in northern Taiwan, where dialysis patients coming from local hemodialysis clinics were treated shunt thrombosis arisen as a complication of dialysis. In those patients, there were two types of shunt thrombosis observed: "acute" and "non-acute" thrombosis. The two types of thrombosis may alternatively recur in a dialysis patient. Figure 1 depicts several observed patterns of recurrence of shunt thrombosis in the study. The researchers in this study are mainly interested in the occurrence of the acute shunt thrombosis, since it would lead to a more aggressive treatment such as surgery. In contrast, a non-acute thrombosis can be handled by a simpler non-invasive treatment. Identification of factors influencing the acute shunt thrombosis in dialysis patients is the major goal of this study.

In the dialysis data mentioned above, the two types of recurrent events, namely the acute and non-acute shunt thrombosis recurrences, can be viewed as bivariate (multivariate) recurrent event data. The data can thus be analyzed by existing methods for multivariate recurrent events data (e.g., Spiekerman and Lin [14]; Cai and Schaubel

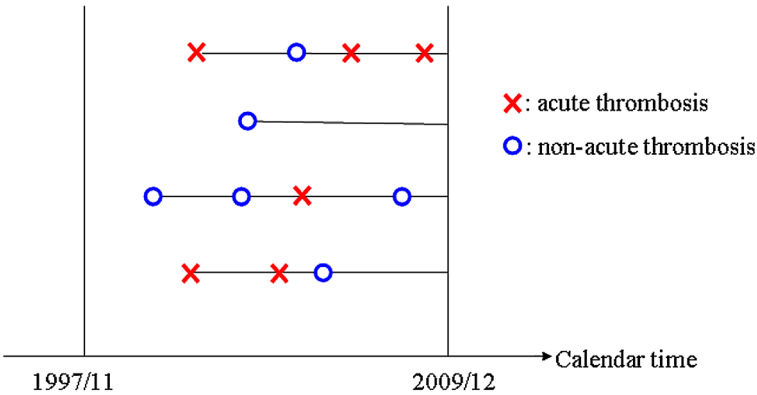

Fig. 1 Possible patterns of recurrences of shunt thrombosis in dialysis patients

[2]; Schaubel and Cai [12]), which are usually based on cause-specific hazards models. However, these methods usually require modeling assumptions for all types of the events considered. In the shunt thrombosis study, however, such assumptions are unnecessarily strong since we are only interested in one of the multiple types of the events, i.e., the acute thrombosis. To relax this limitation, we take an alternative view of the data and pursue a more flexible analysis framework.

Note that, at each time point, each patient may experience at most one of the two types of thrombosis. This implies that we may view the two thrombosis events as two competing risk events. Analysis of competing risks has been a very active area in biomedical practice, and there have been very flexible analysis tools in this area, including the subdistribution hazards regression model proposed by Fine and Gray [6], the mixture regression model by Fine [5], and the binomial regression model by Scheike et al. [13]. These methods, unlike the models based on cause-specific hazards, allow us to specify models only for the event types of interest, and to leave models for other event types fully unspecified. In particular, compared to the subdistribution hazards regression of Fine and Gray [6] and the mixture regression of Fine [5], the binomial regression approach provides direct and flexible estimation of the covariate effects on the overall or cumulative risk for the event of interest, including the time-varying covariate effects (Scheike et al. [13]). In addition, in the shunt thrombosis study, physicians are most interested in factors affecting the cumulative risk of the occurrence of the acute shunt thrombosis reflected in the gap time from the previous thrombosis event to the next one that is of the acute type. We therefore extend the binomial regression model in Scheike et al. [13] from the non-recurrent event analysis to the recurrent event analysis based on gap time, and apply the extended method to the shunt thrombosis data.

This paper is organized as follows. After introducing the motivation and background of this work, in Sect. 2 we describe the data and the model considered. The estimation procedure is presented in detail in Sect. 3. Section 4 reports results from a series of simulations under various scenarios. The application of the proposed method to the shunt thrombosis data motivating the current work is provided in Sect. 5. Section 6 contains some concluding remarks.

2 Data and Model

Suppose there are K types of competing risk events under consideration. Without loss of generality, set $K = 2$, which is the case in the shunt thrombosis study. Let Y_j be the time to the jth occurrence of the event measured from enrollment, and $\Delta_j \in \{1, 2\}$ denote the type of the jth event. In the shunt thrombosis data, $\Delta_j = 1$ or 2 means that the patient suffers from the acute or non-acute thrombosis at the jth event, respectively. Denote by $\mathbf{Z}$ a p-dimensional time-independent covariate vector with the first element being 1. Let C be the time between enrollment and the censoring event due to drop-out or the end of study. We assume that C is completely observable and is independent of Y_j's and $\mathbf{Z}$. Let $\tilde{Y}_j = Y_j \wedge C$ and $\tilde{\Delta}_j = \Delta_j I(Y_j < C)$ for $1 \leq j \leq M$, where M satisfies $Y_{M-1} < C < Y_M$. Consequently, a total of $M - 1$ events are observed exactly over the study period. For $j \geq M$, set $\tilde{\Delta}_j = 0$.

Let $Y_0 = 0$, $T_j = Y_j - Y_{j-1}$ denote the gap time between the $(j-1)$th and jth events for $j \geq 1$. Owing to the existence of the censoring event, T_j may not be observable. What are observable for the gap times are $\tilde{T}_j = \tilde{Y}_j - \tilde{Y}_{j-1}$ for $1 \leq j \leq M$. For $k = 1, 2$, let $N_j^{(k)}(t) = I(T_j \leq t, \Delta_j = k)$ be the underlying counting process for the jth event of type k; here and in the sequel, unless otherwise stated, the time index t is defined in the gap time scale, namely t is reset to 0 at the occurrence of each of the events.

Consider the cumulative incidence function (CIF), also termed the subdistribution of the gap time T_j for type-k competing risk event ($k \in \{1, 2\}$):

$$F_j^{(k)}(t) = E(N_j^{(k)}(t)) = \Pr(T_j \leq t, \Delta_j = k).$$

Suppose that the conditional CIF of the gap time T_j given $\mathbf{Z}$ follows a transformation model:

$$F_j^{(k)}(t|\mathbf{Z}) = g^{(k)}\left\{\mathbf{Z}'\beta_j^{(k)}(t)\right\}, \tag{1}$$

where $g^{(k)}(\cdot)$ is a known monotone link function, $\beta_j^{(k)}(t)$ is the vector of time-varying effects of $\mathbf{Z}$ on the cumulative incidence function of T_j at time t, and a prime denotes transposition. Although $g^{(k)}\{\mathbf{Z}'\beta_j^{(k)}(t)\}$ is required to be nondecreasing in t by definition, $\beta_j^{(k)}(t)$ is not restricted to be nondecreasing in model (1). Note that in (1) we have made the simplification that the link function is the same for each recurrence of competing risk k. Although the proposed procedure will be described under such a simplified setting, its extension to the general setting where the link function can vary with j is obvious.

The model (1) is a generalization of the model in Scheike et al. [13] from non-recurrent event analysis to recurrent event analysis under competing risks. Note that the CIF $F_j^{(k)}(\cdot)$ can be viewed as a subdistribution function for the improper random variable $T_j^* = T_j \times I(\Delta_j = k) + \infty \times I(\Delta_j \neq k)$. In our analysis of the shunt thrombosis data, we adopt model (1) for directly modeling the CIF conditional on the covariates of interest, and propose an estimating equation-based approach for inference on model (1), as detailed in the next section.

Here we introduce several special cases for model (1). If $g^{(k)}(x) = 1 - \exp(-x)$, then $F_j^{(k)}(t|\mathbf{Z}) = 1 - \exp\{-\mathbf{Z}'\beta_j^{(k)}(t)\}$, implying an additive cumulative subdistribution hazard model, namely $-\log\{1 - F_j^{(k)}(t|\mathbf{Z})\} = \mathbf{Z}'\beta_j^{(k)}(t)$. If $g^{(k)}(x) = 1 - \exp\{-\exp(x)\}$, then we have a multiplicative cumulative subdistribution hazard model, namely $-\log\{1 - F_j^{(k)}(t|\mathbf{Z})\} = \exp\{\mathbf{Z}'\beta_j^{(k)}(t)\}$. If $g^{(k)}(x) = \exp(x)/\{1 + \exp(x)\}$, (1) leads to proportional subdistribution odds model given as $\log\left\{\frac{F_j^{(k)}(t|\mathbf{Z})}{1-F_j^{(k)}(t|\mathbf{Z})}\right\} = \mathbf{Z}'\beta_j^{(k)}(t)$.

The observed data can be written as $(\tilde{Y}_{ij}, \tilde{\Delta}_{ij}, \mathbf{Z}_i, C_i, M_i; j = 1, \ldots, M_i, i = 1, \ldots, n)$, which are n independent and identically distributed replicates of $(\tilde{Y}_j, \tilde{\Delta}_j, \mathbf{Z}, C, M; j = 1, \ldots, M)$. The corresponding gap times are defined as $\tilde{T}_{ij} = \tilde{Y}_{ij} - \tilde{Y}_{i(j-1)}$.

3 Parameter Estimation and Model Assessment

In this section, we apply the inverse probability of censoring weighting (IPCW) technique to establish estimating equations for model (1). For some $k \neq 0$, define the counting process $\tilde{N}_{ij}^{(k)}(t) = I(\tilde{T}_{ij} \leq t, \tilde{\Delta}_{ij} = k)$ in terms of the observed gap times. Observe that for $k \neq 0$, $\tilde{N}_{ij}^{(k)}(t) = I(T_{ij} \leq t, \Delta_{ij} = k, C_i > Y_{ij}) = N_{ij}^{(k)}(t) I(C_i > Y_{ij})$. Let $G(s) = \Pr(C > s)$ be the survival function of the censoring time C with the calender time used as the time index. By applying double expectation, we have for $k \neq 0$,

$$
\begin{aligned}
& E\left\{ \frac{\tilde{N}_j^{(k)}(t)}{G(\tilde{Y}_j)} \middle| \mathbf{Z} \right\} \\
&= E\left[E\left\{ \frac{N_j^{(k)}(t) I(C > Y_j)}{G(\tilde{Y}_j)} \middle| Y_j, \mathbf{Z} \right\} \middle| \mathbf{Z} \right] \\
&= E\left[E\left\{ \frac{N_j^{(k)}(t) I(C > Y_j)}{G(Y_j)} \middle| Y_j, \mathbf{Z} \right\} \middle| \mathbf{Z} \right] \\
&= F_j^{(k)}(t) = g^{(k)}\left\{ \mathbf{Z}' \beta_j^{(k)}(t) \right\},
\end{aligned}
$$

motivating us to construct the estimating equation of the form:

$$
\sum_{i=1}^{n} \mathbf{Z}_i \left[\frac{\tilde{N}_{ij}^{(k)}(t)}{\widehat{G}(\tilde{Y}_{ij})} - g^{(k)}\left\{ \mathbf{Z}_i' \beta_j^{(k)}(t) \right\} \right] = 0, \tag{2}
$$

where $\widehat{G}(s)$ is the empirical estimate of the survival function $G(s)$ of the censoring time C based on the data $(C_i; i = 1, \ldots, n)$ (recall that C is always observable in our setting). For obtaining the estimate $\hat{\beta}_j^{(k)}(t)$, it suffices to solve (2) only at the observed gap times $\tilde{T}_{ij}$ with $\tilde{\Delta}_{ij} = k$ for $i = 1, \ldots, n$.

Note that the estimating equation (2) is based on the assumption that the censoring distribution G is independent of the covariates. When G does depend on covariates, it may be estimated by some regression model or the empirical estimates stratified on discretized covariate values.

Let $\hat{\beta}_j^{(k)}(t)$ be the solution of $\beta_j^{(k)}(t)$ to (2). Large sample theory of $\hat{\beta}_j^{(k)}(t)$ can be obtained analogously to that provided in Scheike et al. [13]. In particular, $n^{1/2}\{\hat{\beta}_j^{(k)}(t) - \beta_j^{(k)}(t)\}$ converges to a zero-mean Gaussian process in $\mathcal{B} = (l^\infty[a, \tau])^p$, where $l^\infty[a, \tau]$ is the set of uniformly bounded functions on $[a, \tau]$ with $g^{(k)}(a)$ and $G(\tau)$ bounded above from 0 for almost every covariate value. The asymptotic variance matrix of $n^{1/2}\{\hat{\beta}_j^{(k)}(t) - \beta_j^{(k)}(t)\}$ can be estimated by

$$
n \left(\hat{H}_j^{(k)}(t) \right)^{-1} \left\{ \sum_{i=1}^{n} U_{ij}^{(k)}(t) U_{ij}^{(k)'}(t) \right\} \left(\hat{H}_j^{(k)}(t) \right)^{-1}
$$

evaluated at the estimated parameters, where

$$\begin{aligned}\hat{H}_j^{(k)}(t) &= \frac{\partial}{\partial \beta_j^{(k)}(t)} \sum_{i=1}^{n} \mathbf{Z}_i \left[\frac{\tilde{N}_{ij}^{(k)}(t)}{\widehat{G}(\tilde{Y}_{ij})} - g^{(k)} \left\{ \mathbf{Z}_i' \beta_j^{(k)}(t) \right\} \right] \\ &= -\sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i' \dot{g}^{(k)} \left\{ \mathbf{Z}_i' \beta_j^{(k)}(t) \right\}\end{aligned}$$

with $\dot{g}^{(k)}(x) = \frac{dg^{(k)}(x)}{dx}$, and

$$U_{ij}^{(k)}(t) = \mathbf{Z}_i \left[\frac{\tilde{N}_{ij}^{(k)}(t)}{\widehat{G}(\tilde{Y}_{ij})} - g^{(k)} \left\{ \mathbf{Z}_i' \beta_j^{(k)}(t) \right\} \right] - \frac{1}{n} U_{*j}^{(k)}(t) h_i(\tilde{Y}_{ij})$$

with

$$U_{*j}^{(k)}(t) = \sum_{i=1}^{n} \mathbf{Z}_i \frac{\tilde{N}_{ij}^{(k)}(t)}{\widehat{G}^2(\tilde{Y}_{ij})}, \ h_i(y) = I(C_i \geq y) - \sum_{\ell=1}^{n} I(C_\ell \geq y)/n.$$

Details of the proof are relegated to the Supplementary Material.

Our proposal, like the original binomial modeling approach, estimates the cumulative incidence regression model at each event (gap) time. In the terminology of recurrent event data analysis, the modeling framework we adopt belongs to the so-called marginal model, which aims at assessing factors influencing each occurrence of the event of interest, without reference to any other occurrences of all types of the events. When the inference related to multiple gap times is to be made, for example, when a confidence interval is to be set for $\beta_{j_1}^{(k)}(t_1) - \beta_{j_2}^{(k)}(t_2)$, the difference between coefficient vectors at gap times t_1 and t_2 respectively for the j_1th and j_2th events, one can simply apply the robust sandwich-type variance estimator of $\hat{\beta}_{j_1}^{(k)}(t_1) - \hat{\beta}_{j_2}^{(k)}(t_2)$ given as

$$\left(\hat{H}_{j_1}^{(k)}(t_1) \right)^{-1} \left\{ \sum_{i=1}^{n} U_{i\,j_1}(t_1) U_{i\,j_2}'(t_2) \right\} \left(\hat{H}_{j_2}^{(k)}(t_2) \right)^{-1}$$

to account for the dependence between the gap times, as in the popular generalized estimating equation (GEE)-based procedure for general correlated data and multivariate failure time data (Wei et al. [18]).

To assess the adequacy of a specified model $g^{(k)}\{\mathbf{Z}'\beta_j^{(k)}(t)\}$ for the CIF $F_j^{(k)}(t|\mathbf{Z})$ of the jth event with type k, let $\mathcal{T}_j$ be the set of observed time points where the jth event takes place, and consider a model assessment criterion based on the inverse probability weighting Brier score given by

$$\frac{1}{n}\sum_{t\in\mathcal{T}_j}\sum_{i=1}^{n}\left[\frac{\tilde{N}_{ij}^{(k)}(t)}{\widehat{G}(\tilde{Y}_{ij})}-g^{(k)}\left\{\mathbf{Z}_i'\beta_j^{(k)}(t)\right\}\right]^2$$

which is equivalent to the average of the squared errors $\tilde{N}_{ij}^{(k)}(t)/\widehat{G}(\tilde{Y}_{ij})-g^{(k)}\{\mathbf{Z}_i'\beta_j^{(k)}(t)\}$ over all observed times for the jth event of type k, to compare different subdistribution models such as the proportional subdistribution hazards model or the proportional odds model. The use of this type of model assessment measure is based on the fact that the Brier score is a well-known measure for accuracy of probability forecasts (see for example Wen and Chen [19]), which is suitable for our framework of cumulative incidence function analysis. The inverse probability weighting is applied when performing such model assessment to account for censoring.

4 Simulation Study

Simulations are performed to illustrate the finite sample performances of the proposed analysis under several scenarios, where there are two types of competing risk events and each of them can recur. To be focused, we consider only estimation results for the regression parameters $\hat{\beta}_j^{(1)}(t)$, i.e., the coefficient function corresponding to type-1 competing risk. Both the multiplicative cumulative subdistribution hazard model, i.e., $g^{(1)}(x)=1-\exp\{-\exp(x)\}$, and the proportional subdistribution odds model, i.e., $g^{(1)}(x)=\exp(x)/\{1+\exp(x)\}$, are considered.

A three-dimensional covariate vector $\mathbf{Z}$ is considered, where the first element is 1, the second element is generated from a normal distribution with mean 1 and variance 0.49 and is truncated to have a range of $(0,2)$, and the third element is a bernoulli distribution with success probability 0.5. The corresponding regression parameters for the first occurrence of the event, $\beta_1^{(1)}(t)=(\beta_{10}^{(1)}(t),\beta_{11}^{(1)}(t),\beta_{12}^{(1)}(t))'$, is set to:

$$\begin{aligned}
\beta_{10}^{(1)}(t)&=h(\gamma\cdot\{1-\exp(-1.25t)\}),\\
\beta_{11}^{(1)}(t)&=0.5(1-\exp(-t)),\\
\beta_{12}^{(1)}(t)&=0,
\end{aligned}$$

where $h(\cdot)$ is the inverse function of $g^{(1)}(\cdot)$, and $\gamma=0.5$, 0.6 or 0.8 is the proportion of type-1 competing risk at $t=\infty$. The coefficient function $\beta_2^{(1)}(t)$ for the second occurrence of the type-1 competing risk is set to

$$\begin{aligned}
\beta_{20}^{(1)}(t)&=h(\gamma\cdot\{1-\exp(-1.25t)\}),\\
\beta_{21}^{(1)}(t)&=\begin{cases}0, & \text{when}\quad \gamma=0.8,\\ \frac{\exp(3t)-1}{2(1+\exp(3t))}, & \text{when}\quad \gamma=0.6,\\ \frac{0.5t}{t+1}, & \text{when}\quad \gamma=0.5,\end{cases}\\
\beta_{22}^{(1)}(t)&=0.
\end{aligned}$$

In all the settings considered, we examine the bias, simulation standard deviation (SD), the mean of the estimated standard errors (SE), and the empirical coverage probabilities (CP) of the point-wise Wald-type 95% confidence interval for the regression coefficient functions at four specified time points, which correspond to time points at which the survival function for the second gap time $\Pr(T_2 > t|\mathbf{Z} = (1, 1, 0.5)') = 0.8, 0.6, 0.4$, and 0.2. An independent censoring time is generated from a uniform distribution in (0, A), where A is chosen such that the censoring rates for (T_1, T_2) from the type-1 competing risk is about (15, 30 %) or (20, 40 %).

Following Cheng and Fine [3], for the type-1 competing risk, the event times are assumed to have a positive correlation, and are conditionally independent given the frailty variable W which is generated from a gamma distribution with both the mean and variance being 2. Specifically, given the specified type-1 CIF model of $F_j^{(1)}(t|\mathbf{Z})$, the simulated covariate value $\mathbf{z}$ and the frailty variable value w, the conditional type-1 CIF given $W = w$ is

$$F_j^{(1)}(t|w, \mathbf{z}) = \Pr(T_j \le t, \Delta_j = 1|\mathbf{Z} = \mathbf{z}, W = w) = \exp\left\{ - w \cdot q\left\{F_j^{(1)}(t|\mathbf{z})\right\}\right\}$$

with $q(\cdot)$ the inverse Laplace transformation of W. The event times from the type-2 competing risk are assumed to be marginally independent. Let c the simulated value of the censoring time. The gap times (T_1, T_2) for the first two events of a subject are then generated by the following procedure.

Step 1: For $j = 1, 2$, a value u is generated from a uniform distribution in (0,1). When $u < F_j^{(1)}(\infty|w, \mathbf{z})$, $\Delta_j = 1$ and T_j is generated from $F_j^{(1)}(t|w, \mathbf{z})$. Otherwise, $\Delta_j = 2$ and T_j should be generated from the conditional distribution $\Pr(T_j \le t|Z = z, \Delta_j = 2)$, instead of CIF. Set $Y_j = \sum_{l=1}^{j} T_l$.

Step 2: If $Y_j \le c$, set $\tilde{Y}_j = Y_j$, $\tilde{T}_j = T_j$ and $\tilde{\Delta}_j = \Delta_j$. Otherwise, set $M = j$, $\tilde{Y}_M = c$, $\tilde{T}_M = c - Y_{M-1}$, and $\tilde{\Delta}_M = 0$.

Under the scenario mentioned above and the CIF of the type-1 competing risk is given by a multiplicative cumulative subdistribution hazard model for the first two events, Table 1 shows the results for $F_{j0}^{(1)}(t) = F_j^{(1)}(t|\mathbf{Z} = (1, 0, 0)') = g^{(1)}(\beta_{j0}^{(1)}(t))$, and for each elements of $\beta_j^{(1)}(t)$ at four given time points, $j = 1, 2$. To save space, we only report results with a sample size of 200 and censoring rates (15, 30 %) in the first two events; results for smaller sample size $n = 100$ and higher censoring rates (20, 40 %) in the first two events have similar performance patterns and are displayed in the Supplementary Material. We can see that in both of the two scenarios, the proposed estimates are nearly unbiased, and the simulation standard deviations are close to the means of the estimated standard errors. The resulting coverage probabilities achieve the nominal level 0.95 well.

Similarly, the proportional subdistribution odds model is applied for $g^{(1)}(\cdot)$ in (1). We report in Table 2 the results and see the nice performances again for the proposed estimates.

Table 1 Simulation results for $g^{(1)}(x) = 1 - \exp\{-\exp(x)\}$

γ	t	$F_{10}^{(1)}(t)$					$\beta_{10}^{(1)}(t)$					$\beta_{11}^{(1)}(t)$					$\beta_{12}^{(1)}(t)$				
		True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP
Scenario 1: the regression parameters at event 1																					
0.8	0.175	0.157	0.002	0.053	0.055	0.950	−1.767	−0.048	0.389	0.393	0.952	0.080	0.015	0.291	0.293	0.958	0	−0.015	0.360	0.358	0.942
	0.400	0.315	0.005	0.072	0.071	0.946	−0.972	−0.004	0.285	0.276	0.954	0.165	0.000	0.206	0.205	0.960	0	−0.006	0.239	0.248	0.956
	0.718	0.474	0.004	0.077	0.078	0.964	−0.442	0.001	0.233	0.236	0.962	0.256	0.005	0.176	0.177	0.958	0	−0.008	0.213	0.212	0.954
	1.263	0.635	−0.002	0.082	0.086	0.966	0.008	−0.005	0.229	0.239	0.968	0.359	0.019	0.182	0.193	0.972	0	−0.007	0.229	0.225	0.948
0.6	0.195	0.130	0.004	0.053	0.051	0.942	−1.974	−0.050	0.452	0.431	0.952	0.088	0.012	0.338	0.319	0.944	0	−0.021	0.414	0.396	0.958
	0.411	0.241	0.003	0.067	0.064	0.932	−1.288	−0.023	0.333	0.314	0.928	0.169	0.002	0.238	0.232	0.944	0	−0.005	0.293	0.282	0.936
	0.687	0.346	0.001	0.076	0.072	0.932	−0.857	−0.018	0.284	0.267	0.934	0.248	0.001	0.200	0.197	0.952	0	0.000	0.255	0.238	0.930
	1.152	0.458	0.002	0.083	0.080	0.954	−0.491	−0.005	0.255	0.246	0.932	0.342	−0.005	0.189	0.185	0.932	0	0.007	0.226	0.221	0.944
0.5	0.204	0.112	0.001	0.047	0.047	0.936	−2.126	−0.087	0.483	0.466	0.938	0.092	0.027	0.353	0.345	0.944	0	0.019	0.450	0.425	0.962
	0.459	0.218	0.003	0.062	0.062	0.936	−1.402	−0.021	0.328	0.330	0.942	0.184	0.006	0.250	0.242	0.946	0	−0.001	0.303	0.293	0.934
	0.820	0.321	−0.002	0.066	0.071	0.964	−0.951	−0.025	0.255	0.277	0.966	0.280	0.016	0.192	0.203	0.966	0	0.005	0.251	0.244	0.942
	1.484	0.422	0.002	0.074	0.080	0.974	−0.602	−0.007	0.237	0.257	0.966	0.387	0.007	0.184	0.192	0.958	0	0.000	0.223	0.229	0.952
Scenario 2: the regression parameters at event 2																					
0.8	0.175	0.157	0.006	0.064	0.061	0.933	−1.767	−0.042	0.458	0.428	0.929	0	0.009	0.345	0.322	0.924	0	0.006	0.398	0.397	0.969
	0.400	0.315	0.005	0.088	0.081	0.929	−0.972	−0.019	0.350	0.321	0.940	0	0.013	0.261	0.243	0.933	0	−0.003	0.285	0.295	0.971
	0.718	0.474	0.002	0.103	0.095	0.944	−0.442	−0.010	0.311	0.290	0.942	0	0.013	0.233	0.220	0.949	0	−0.004	0.277	0.266	0.958
	1.263	0.635	−0.001	0.105	0.105	0.951	0.008	−0.002	0.300	0.300	0.960	0	0.003	0.230	0.227	0.966	0	0.004	0.281	0.274	0.951

Table 1 continued

γ	t	$F_{10}^{(1)}(t)$					$\beta_{10}^{(1)}(t)$					$\beta_{11}^{(1)}(t)$					$\beta_{12}^{(1)}(t)$				
		True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP
0.6	0.195	0.130	0.006	0.058	0.055	0.931	−1.974	−0.051	0.496	0.464	0.940	0.142	−0.008	0.348	0.341	0.960	0	−0.012	0.421	0.421	0.967
	0.411	0.241	0.007	0.077	0.072	0.947	−1.288	−0.011	0.361	0.347	0.942	0.274	−0.009	0.268	0.255	0.953	0	−0.011	0.307	0.309	0.958
	0.687	0.346	0.006	0.087	0.086	0.944	−0.857	−0.007	0.318	0.313	0.944	0.387	−0.004	0.232	0.234	0.956	0	−0.006	0.280	0.278	0.949
	1.152	0.458	0.003	0.100	0.101	0.955	−0.491	−0.010	0.310	0.315	0.966	0.469	0.004	0.252	0.254	0.975	0	−0.002	0.295	0.292	0.949
0.5	0.204	0.112	0.003	0.052	0.052	0.944	−2.126	−0.084	0.528	0.510	0.953	0.085	−0.011	0.389	0.379	0.949	0	0.035	0.474	0.470	0.966
	0.459	0.218	0.004	0.073	0.070	0.946	−1.402	−0.030	0.385	0.373	0.947	0.157	−0.014	0.288	0.278	0.936	0	0.028	0.320	0.338	0.962
	0.820	0.321	0.002	0.083	0.084	0.949	−0.951	−0.024	0.320	0.327	0.956	0.225	−0.006	0.249	0.245	0.947	0	0.021	0.283	0.296	0.962
	1.484	0.422	0.004	0.102	0.099	0.958	−0.602	−0.010	0.326	0.322	0.956	0.299	−0.007	0.251	0.247	0.953	0	0.020	0.300	0.295	0.955

The statistics are based on 500 replications with $n = 200$ and the censoring rates are around (15, 30 %) for (T_1, T_2); SD is the simulation sample standard deviation, SE is the mean of the estimated standard errors, CP is coverage probability of 95 % confidence interval based on asymptotic normality

Table 2 Simulation results for $g^{(1)}(x) = \exp(x)/\{1 + \exp(x)\}$

γ	t	$F_{10}^{(1)}(t)$					$\beta_{10}^{(1)}(t)$					$\beta_{11}^{(1)}(t)$					$\beta_{12}^{(1)}(t)$				
		True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP
Scenario 1: the regression parameters at event 1																					
0.8	0.175	0.157	0.002	0.057	0.055	0.938	−1.680	−0.055	0.457	0.429	0.936	0.080	0.018	0.340	0.320	0.944	0	0.010	0.374	0.392	0.960
	0.400	0.315	0.003	0.073	0.072	0.940	−0.777	−0.007	0.345	0.340	0.962	0.165	0.003	0.259	0.257	0.942	0	0.002	0.297	0.310	0.972
	0.718	0.474	0.000	0.080	0.082	0.956	−0.104	−0.002	0.328	0.336	0.960	0.256	0.024	0.261	0.259	0.954	0	−0.025	0.312	0.311	0.952
	1.263	0.635	0.000	0.087	0.089	0.958	0.554	0.021	0.391	0.398	0.962	0.359	0.017	0.313	0.318	0.956	0	−0.045	0.379	0.380	0.952
0.6	0.181	0.121	0.004	0.051	0.050	0.944	−1.979	−0.044	0.481	0.480	0.944	0.083	0.004	0.359	0.360	0.958	0	−0.018	0.430	0.439	0.970
	0.388	0.231	0.005	0.064	0.065	0.956	−1.204	−0.008	0.371	0.372	0.968	0.161	0.006	0.290	0.280	0.950	0	−0.025	0.346	0.337	0.942
	0.663	0.338	0.007	0.074	0.075	0.960	−0.672	0.012	0.338	0.340	0.962	0.242	0.002	0.259	0.257	0.958	0	−0.025	0.310	0.310	0.940
	1.138	0.455	0.002	0.082	0.084	0.966	−0.179	0.003	0.339	0.346	0.968	0.340	0.009	0.266	0.268	0.946	0	−0.020	0.308	0.322	0.968
0.5	0.189	0.105	0.000	0.049	0.045	0.922	−2.139	−0.106	0.550	0.514	0.944	0.086	0.045	0.388	0.379	0.926	0	−0.011	0.474	0.471	0.96
	0.426	0.206	−0.003	0.060	0.061	0.946	−1.346	−0.061	0.378	0.387	0.962	0.173	0.029	0.279	0.289	0.956	0	−0.007	0.340	0.3503	0.962
	0.761	0.307	−0.003	0.069	0.072	0.964	−0.815	−0.035	0.330	0.349	0.964	0.266	0.015	0.254	0.263	0.964	0	−0.009	0.318	0.317	0.952
	1.362	0.409	−0.001	0.084	0.082	0.942	−0.368	−0.014	0.356	0.351	0.950	0.372	0.024	0.284	0.272	0.948	0	−0.031	0.347	0.325	0.946
Scenario 2: the regression parameters at event 2																					
0.8	0.175	0.157	0.004	0.064	0.060	0.944	−1.680	−0.053	0.494	0.474	0.946	0	0.005	0.375	0.358	0.953	0	−0.019	0.434	0.437	0.949
	0.400	0.315	−0.002	0.086	0.081	0.929	−0.777	−0.037	0.415	0.393	0.938	0	0.008	0.293	0.297	0.953	0	0.002	0.368	0.358	0.956
	0.718	0.474	−0.006	0.097	0.095	0.947	−0.104	−0.029	0.407	0.398	0.958	0	0.011	0.295	0.301	0.960	0	0.015	0.365	0.363	0.951
	1.263	0.635	−0.015	0.110	0.107	0.946	0.554	−0.034	0.511	0.498	0.949	0	0.033	0.398	0.381	0.962	0	0.014	0.462	0.453	0.946

Table 2 continued

γ	t	$F_{10}^{(1)}(t)$					$\beta_{10}^{(1)}(t)$					$\beta_{11}^{(1)}(t)$					$\beta_{12}^{(1)}(t)$				
		True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP	True	Bias	SD	SE	CP
0.6	0.181	0.121	0.005	0.055	0.054	0.947	−1.979	−0.051	0.540	0.517	0.956	0.132	0.007	0.405	0.380	0.960	0	−0.010	0.488	0.470	0.946
	0.388	0.231	0.002	0.072	0.071	0.944	−1.204	−0.034	0.426	0.416	0.951	0.262	0.016	0.334	0.309	0.936	0	0.000	0.357	0.374	0.966
	0.663	0.338	−0.003	0.090	0.086	0.940	−0.672	−0.042	0.417	0.403	0.951	0.380	0.023	0.330	0.310	0.931	0	0.008	0.343	0.369	0.978
	1.138	0.455	−0.010	0.107	0.104	0.946	−0.179	−0.047	0.454	0.443	0.969	0.468	0.044	0.365	0.358	0.958	0	0.002	0.399	0.423	0.971
0.5	0.189	0.105	0.002	0.052	0.050	0.949	−2.139	−0.102	0.597	0.560	0.953	0.080	0.015	0.439	0.415	0.942	0	0.002	0.524	0.515	0.960
	0.426	0.206	0.000	0.073	0.069	0.942	−1.346	−0.059	0.463	0.440	0.947	0.149	0.016	0.329	0.330	0.958	0	−0.001	0.414	0.399	0.944
	0.761	0.307	0.000	0.087	0.084	0.929	−0.815	−0.036	0.425	0.411	0.953	0.216	0.019	0.315	0.311	0.947	0	−0.015	0.365	0.374	0.958
	1.362	0.409	−0.009	0.099	0.099	0.947	−0.368	−0.057	0.432	0.434	0.964	0.288	0.032	0.338	0.338	0.966	0	0.000	0.397	0.402	0.960

The statistics are based on 500 replications with $n = 200$, and the censoring rates are around $(15, 30\,\%)$ for (T_1, T_2); SD is the simulation sample standard deviation, SE is the mean of the estimated standard errors, CP is coverage probability of 95 % confidence interval based on asymptotic normality

5 Data Analysis

We apply the proposed analysis based on the cumulative incidence probability model (1) to the shunt thrombosis data, which was collected by a large hospital in northern Taiwan. The follow-up time is from November, 1997 to December, 2009, and 2886 dialysis patients with kidney diseases participated in this clinical study. As pointed out in Introduction, two types of shunt thrombosis, designated as the "acute" and "non-acute" thrombosis, may recur during the hemodialysis treatment. In the following data analysis, we are interested in the "acute" type of thrombosis and its cumulative incidence probabilities in the first two events are analyzed via model (1) with $j = 1, 2$. Eight covariates besides the intercept are included in model (1) whose effects on the jth gap time to the occurrence of acute shunt thrombosis are to be assessed; see Table 3 for definitions, counts and means (standard deviations) for these covariates across the first five events stratified by types of thrombosis. For each individual, Y_0 is set to be the calendar time of enrollment as of November 1st, 1997, and C is set to be the last date of 2009. After deleting records with missing information, the data set used in our analysis consists of 2779 subjects.

We fit to the shunt thrombosis data the multiplicative cumulative subdistribution hazard model, i.e., $g^{(1)}(x) = 1 - \exp\{-\exp(x)\}$ in model (1), for the incidence of acute shunt thrombosis. The estimated coefficient functions for the first event are displayed in Fig. 2. We see that, after accounting for the point-wise confidence interval, the coefficient function of the covariate "hypertension" is essentially positive over the range of the duration time of the first event except at initial and later time. This result suggests that dialysis patients with hypertension are associated with higher incidence of acute shunt thrombosis relative to patients without hypertension. In addition, the estimated regression coefficient function of the covariate "shunt type" is also significantly above zero over the gap time to the first event except for an initial small time interval. This coincides with the prior knowledge that acute shunt thrombosis is more likely to arise from a shunt of the graft type than a shunt of the natural type. The other covariates, however, do not show significant effects on the incidence of the acute shunt thrombosis over the whole range of the gap time to the first event.

Similarly, Fig. 3 shows the estimated coefficient functions corresponding to the second event. Among the covariates considered, the "shunt type" is the only covariate to have statistically significant time-varying effects on the incidence of acute shunt thrombosis over the whole range of the gap time to the second event except at the initial time. Hypertension, although exhibits significant effects in the gap time to the first event, no longer affects the incidence of acute thrombosis in the gap time to the second event. In further analysis up to the fifth event (results not shown), we find that the shunt type continues to be a significant covariate for acute thrombosis in the first five events, while hypertension is a significant covariate for acute thrombosis only in the first event and its effects on incidence of acute thrombosis diminish in the later events.

We also fit to the data a proportional subdistribution odds model, i.e., $g^{(1)}(x) = \exp(x)/\{1 + \exp(x)\}$ in model (1), for the incidence of acute shunt thrombosis in the first j events ($j = 1, \ldots, 5$). The results are qualitatively equivalent to those from the multiplicative cumulative subdistribution hazard model described above, and hence

Table 3 The count (proportion[a]) and mean (standard deviation) of the eight covariates we considered cross-classified by events in the shunt thrombosis data; CAD means coronary artery disease, and age means age at enrollment

Variable	Category (coding)	Event 1 ($n = 2779$)		Event 2 ($n = 1641$)		Event 3 ($n = 1112$)		Event 4 ($n = 823$)		Event 5 ($n = 641$)	
		Acute	Non-acute	Acute	Non-acute	Acute	Non-acute	Acute	Non-acute	Acute	Non-acute
Gender	Female (0)	231 (15.63%)	663 (44.86%)	181 (12.25%)	439 (29.70%)	149 (10.08%)	328 (22.19%)	129 (8.73%)	242 (16.37%)	101 (6.83%)	202 (13.67%)
	Male (1)	173 (13.30%)	574 (44.12%)	149 (11.45%)	343 (26.36%)	127 (9.76%)	219 (16.83%)	95 (7.30%)	175 (13.45%)	67 (5.15%)	137 (10.53%)
Hypertension	No (0)	177 (14.27%)	573 (46.21%)	157 (12.66%)	351 (28.31%)	122 (9.84%)	253 (20.40%)	99 (7.98%)	193 (15.56%)	79 (6.37%)	159 (12.82%)
	Yes (1)	227 (14.75%)	664 (43.14%)	173 (11.24%)	431 (28.01%)	154 (10.01%)	294 (19.10%)	125 (8.12%)	224 (14.55%)	89 (5.78%)	180 (11.70%)
Smoking	No (0)	362 (15.10%)	1066 (44.47%)	287 (11.97%)	692 (28.87%)	244 (10.18%)	486 (20.28%)	197 (8.22%)	370 (15.44%)	153 (6.38%)	299 (12.47%)
	Yes (1)	42 (10.99%)	171 (44.76%)	43 (11.26%)	90 (23.56%)	32 (8.38%)	61 (15.97%)	27 (7.07%)	47 (12.30%)	15 (3.93%)	40 (10.47%)
Hyperlipidemia	No (0)	371 (15.01%)	1101 (44.56%)	285 (11.53%)	705 (28.53%)	251 (10.16%)	483 (19.55%)	201 (8.13%)	373 (15.10%)	149 (6.03%)	304 (12.30%)
	Yes (1)	33 (10.71%)	136 (44.16%)	45 (14.61%)	77 (25.00%)	25 (8.12%)	64 (20.78%)	23 (7.47%)	44 (14.29%)	19 (6.17%)	35 (11.36%)
Diabetes	No (0)	215 (14.46%)	669 (44.99%)	180 (12.10%)	420 (28.24%)	143 (9.62%)	305 (20.51%)	131 (8.81%)	228 (15.33%)	96 (6.46%)	180 (12.10%)
	Yes (1)	189 (14.63%)	568 (43.96%)	150 (11.61%)	362 (28.02%)	133 (10.29%)	242 (18.73%)	93 (7.20%)	189 (14.63%)	72 (5.57%)	159 (12.31%)
CAD	No (0)	353 (14.48%)	1085 (44.50%)	292 (11.98%)	684 (28.06%)	239 (9.80%)	482 (19.77%)	201 (8.24%)	370 (15.18%)	145 (5.95%)	300 (12.31%)
	Yes (1)	51 (14.96%)	152 (44.57%)	38 (11.14%)	98 (28.74%)	37 (10.85%)	65 (19.06%)	23 (6.74%)	47 (13.78%)	23 (6.74%)	39 (11.44%)
Shunt type	Natural (0)	263 (12.71%)	882 (42.61%)	214 (10.34%)	514 (24.83%)	166 (8.02%)	350 (16.91%)	132 (6.38%)	253 (12.22%)	99 (4.78%)	193 (9.32%)
	Graft (1)	141 (19.89%)	355 (50.07%)	116 (16.36%)	268 (37.80%)	110 (15.51%)	197 (27.79%)	92 (12.98%)	164 (23.13%)	69 (9.73%)	146 (20.59%)
Age	–	63.84 (12.97)	63.42 (13.50)	63.41 (12.80)	64.02 (13.07)	63.77 (12.48)	64.53 (12.87)	64.73 (12.76)	63.53 (12.66)	63.62 (11.36)	64.32 (12.99)

[a] The proportion is taken with respect to the baseline count in each category

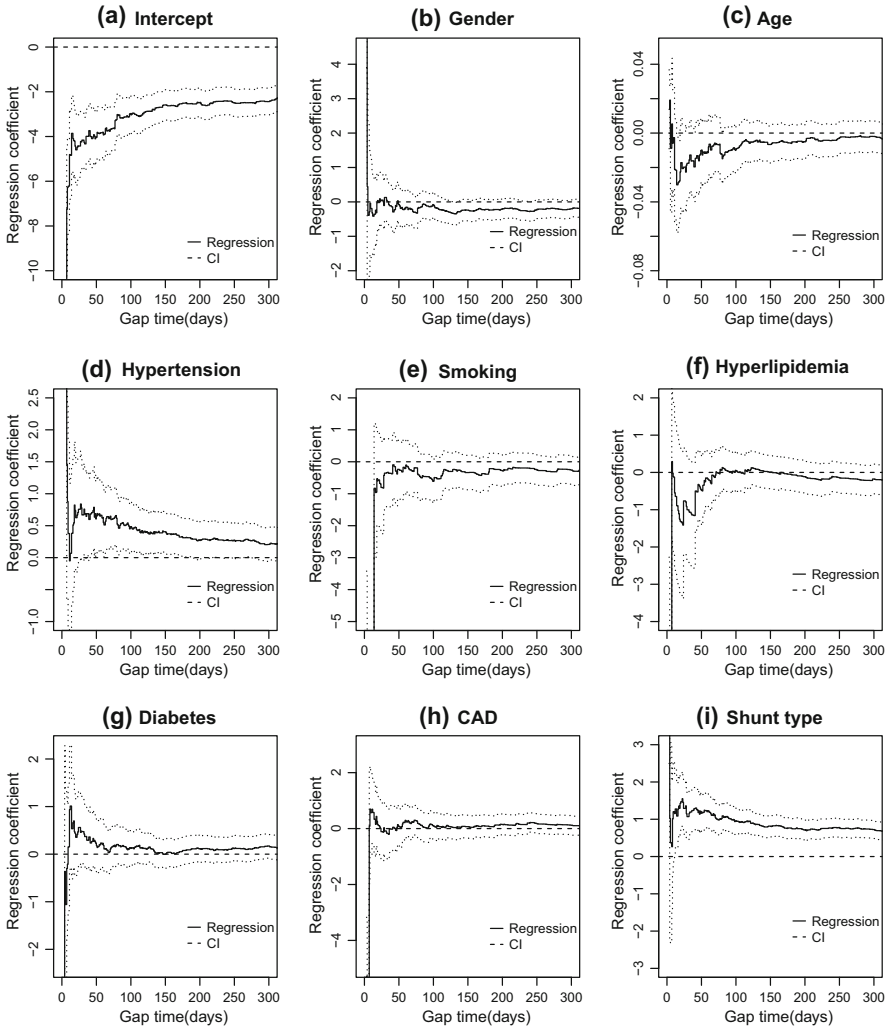

Fig. 2 The estimated regression coefficient functions (and point-wise 95 % confidence intervals) corresponding to acute thrombosis under the multiplicative cumulative subdistribution hazard model at the first event

are omitted here. Additionally, we perform the model assessment based on the inverse probability weighting Brier score, which gives (22.78586, 17.33794) in the first two events for the multiplicative cumulative subdistribution hazard model, respectively. On the other hand, the proportional subdistribution odds model gives slightly large scores (22.78716, 17.33848) in the first two events. So the multiplicative cumulative subdistribution hazard model is a better model in the shunt thrombosis data according to the Brier-based criterion.

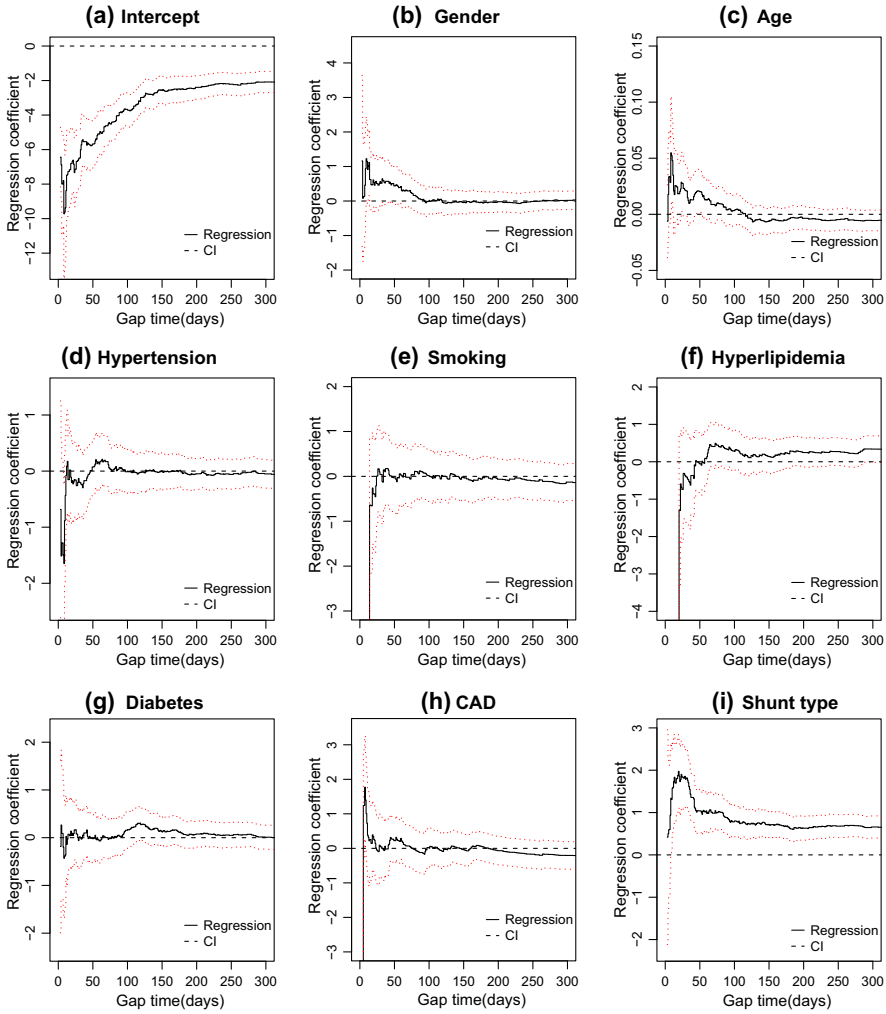

Fig. 3 The estimated regression coefficient functions (and point-wise 95 % confidence intervals) corresponding to acute thrombosis under the multiplicative cumulative subdistribution hazard model at the second event

6 Discussion

In this paper we report an analysis of gap times between repeated occurrences of shunt thrombosis in dialysis patients, where the shunt thrombosis is classified into two types of main interest. Instead of treating the two-type shunt thrombosis as a bivariate recurrent event, we view the "acute" and "non-acute" thrombosis that may occur on a patient at a time point as two competing risk events. This option allows us to adapt the tools for regression analysis of competing risks recently developed to study the effects of certain covariates on the gap time between two shunt thrombosis occurrences. In

particular, we focus on direct regression analysis of the subdistribution or cumulative incidence probability function so that we are able to identify factors influencing the occurrence of a specific type of shunt thrombosis, and to allow for the flexibility of time-varying effects of the factors.

Our analysis of the shunt thrombosis data reveals that, dialysis patients with the graft type of shunt tend to have faster occurrence of acute shunt thrombosis. The subdistribution regression model renders it convenient to perform prediction of the occurrence of acute shunt thrombosis. Hypertension also leads to increased incidence probabilities of acute shunt thrombosis, but only for the initial thrombosis event and not for subsequent recurrences of thrombosis.

The binomial regression model we employ to model the cumulative incidence of recurrent events is semiparametric in that the functional form of the covariate effect is left fully unspecified. The resulting semiparametric analysis is thus adaptive to rather complicated data structures. Fitting such a model to the recurrent event data, we can examine time-varying covariate effects on each of the recurrent events.

One limitation of the proposed method is that it cannot accommodate covariates whose values vary over different events. The extension for such time-dependent covariates seems non-trivial and deserves substantial further research.

References

1. Andersen PK, Gill R (1982) Cox's regression model for counting processes, a large sample study. Ann Stat 10(4):1100–1120
2. Cai J, Schaubel DE (2004) Marginal means/rates models for multiple type recurrent event data. Lifetime Data Anal 10(2):121–138
3. Cheng Y, Fine JP (2012) Cumulative incidence association models for bivariate competing risks data. J R Stat Soc Ser B 74(2):183–202
4. Cook RJ, Lawless JF (2007) The statistical analysis of recurrent events. Springer, New York
5. Fine JP (1999) Analyzing competing risks data with transformation models. J R Stat Soc Ser B 61(4):817–830
6. Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94(446):496–509
7. Huang YJ, Chen YQ (2003) Marginal regression of gaps between recurrent events. Lifetime Data Anal 9(3):293–303
8. Huang X, Liu L (2007) A joint frailty model for survival and gap times between recurrent events. Biometrics 63(2):389–397
9. Lu W (2005) Marginal regression of multivariate event times based on linear transformation models. Lifetime Data Anal 11(3):389–404
10. Prentice RL, Williams BJ, Peterson AV (1981) On the regression analysis of multivariate failure time data. Biometrika 68(2):373–379
11. Schaubel DE, Cai J (2004) Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. Biometrika 91(2):291–303
12. Schaubel DE, Cai J (2006) Rate/mean regression for multiple-sequence recurrent event data with missing event category. Scand J Stat 33(2):191–207
13. Scheike TH, Zhang MJ, Gerds TA (2008) Predicting cumulative incidence probability by direct binomial regression. Biometrika 95(1):205–220
14. Spiekerman CF, Lin DY (1998) Marginal regression models for multivariate failure time data. J Am Stat Assoc 93(443):1164–1175
15. Sun LQ, Park DH, Sun JG (2006) The additive hazards model for recurrent gap times. Stat Sin 16:919–932

16. Zeng D, Lin DY (2006) Efficient estimation of semiparametric transformation models for counting processes. Biometrika 93(3):627–640
17. Wang MC, Qin J, Chiang CT (2001) Analyzing recurrent event data with informative censoring. J Am Stat Assoc 96(455):1057–1065
18. Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. J Am Stat Assoc 84(408):1065–1073
19. Wen CC, Chen YH (2013) Assessing age-at-onset risk factors with incomplete covariate current status data under proportional odds models. Stat Med 32(12):2001–2012