BIOMETRICS 68, 637–647
June 2012

DOI: 10.1111/j.1541-0420.2011.01676.x

# Computationally Efficient Marginal Models for Clustered Recurrent Event Data

**Dandan Liu,[1,*] Douglas E. Schaubel,[2] and John D. Kalbfleisch[2]**

[1]Department of Biostatistics, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, Tennessee 37232, U.S.A.

[2]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

[*]*email:* dandanl@gmail.com

SUMMARY. Large observational databases derived from disease registries and retrospective cohort studies have proven very useful for the study of health services utilization. However, the use of large databases may introduce computational difficulties, particularly when the event of interest is recurrent. In such settings, grouping the recurrent event data into prespecified intervals leads to a flexible event rate model and a data reduction that remedies the computational issues. We propose a possibly stratified marginal proportional rates model with a piecewise-constant baseline event rate for recurrent event data. Both the absence and the presence of a terminal event are considered. Large-sample distributions are derived for the proposed estimators. Simulation studies are conducted under various data configurations, including settings in which the model is misspecified. Guidelines for interval selection are provided and assessed using numerical studies. We then show that the proposed procedures can be carried out using standard statistical software (e.g., `SAS`, `R`). An application based on national hospitalization data for end-stage renal disease patients is provided.

KEY WORDS: Clustered recurrent event data; Interval-grouped data; Large database; Marginal models; Piecewise constant; Proportional rates.

## 1. Introduction

Hospitalizations are generally very costly events. For example, hospital stays represent over one third of total Medicare expenditures for dialysis patients (U.S. Renal Data System, 2006). Quantifying the impact of patient characteristics on the frequency and duration of hospitalization is an essential step toward the controlling of escalating medical costs, and can play an important role in providing cost-effective health care. In addition, assessment of dialysis facility outcomes in terms of hospitalization and comparison with outcomes at the national level can help to enhance a facility's understanding of its quality of care and how it relates to other facilities. Therefore, statistical modeling of hospitalization is needed to estimate and compare hospitalization rates. Because dialysis patients may have multiple hospital admissions, both hospital admissions (reflecting incidence) and hospital days (reflecting prevalence) can be considered as recurrent event data. Moreover, clustering is introduced both through the dependence among patients in the same facility and the correlation of outcomes over time for a given patient.

Many statistical methods have been proposed for recurrent event data (e.g., Andersen and Gill, 1982; Lawless and Nadeau, 1995; Lin et al., 2000). The semiparametric proportional rates model of Lin et al. (2000) is widely used due to the ease of its implementation with standard statistical software such as `SAS` and `R`. The method has been extended to accommodate clustered recurrent event data. For example, Schaubel and Cai (2005) proposed two extensions applicable to clustered recurrent event data. The first assumes a cluster-specific baseline rate function, while the second assumes a common baseline rate function. It should be noted that each of the aforementioned methods requires the observation of each event occurrence time. Each is rank based and, as such, uses the exact event times to order the failure and censoring time to construct the risk sets (and related summations) appropriately.

The analysis that motivated our current work considers the hospitalization experience among U.S. dialysis patients using both national end-stage renal disease (ESRD) registry data and that obtained from the Centers for Medicare and Medicaid Services (CMS). The pertinent analysis file is extremely large because there are over 5000 dialysis facilities in the United States and more than 500,000 ESRD patients receiving dialysis treatment each year. Each dialysis patient may have multiple hospital admissions every year; on average, patients have 1.25 admissions with an average stay of 8 days for each admission. It has been known for some time that standard Cox regression software (e.g., `R`'s `coxph(·)`, `SAS`'s `PROC PHREG`) can be used to fit the proportional rates model of Lin et al. (2000). Specifically, each patient's follow-up is represented by a set of records, one per recurrent event (plus one for the final censoring event). For example, the experience of a patient with events at 4, 7, 9, and censored at time 12 would be represented by four records: (0,4], (4,7], (7,9], and (9,12]; the event indicator would equal 1 for the first three records and 0 for the last record. This has come to be known as the 'counting process' style data structure; e.g., as described in Allison (2010). In our motivating example, the number of days hospitalized is of interest, as opposed to hospital admissions. If one uses the left-truncated data structure just described, it is clear that even moderate-sized data sets can become unduly large if subjects are hospitalized frequently or tend to have long stays. For example, a hospitalization at time 7 with duration 8 days would result in nine separate records: (6,7], (7,8], ..., (13,14]; the event indicator set to 1 for each. Therefore, our use of the U.S. national ESRD and CMS databases will introduce computational difficulties. In settings such as these, the use of a piecewise-constant recurrent event rate model allows for the grouping of the recurrent event data, which leads to a flexible event rate model and a resulting data reduction that ameliorates the computational burden.

The proposed methods involve grouping recurrent events into intervals corresponding to the 'pieces' implied by the assumed piecewise-constant baseline rate function. With respect to interval-grouped event data, several authors have investigated nonparametric methods in estimating the mean and rate functions (e.g., Thall and Lachin, 1988; Sun and Kalbfleisch, 1995; Wellner and Zhang, 2000). However, such methods do not consider covariate effects. Lawless and Zhan (1998) proposed a proportional rates model with a piecewise-constant baseline rate for interval-grouped recurrent event data. The authors developed robust estimation techniques based on generalized estimating equations, without assumptions on the event process. However, such methods assume a common baseline rate function and may not be easily extended to the situation where the baseline is cluster specific, especially when the number of clusters is relatively large. Sun and Wei (2000) proposed semiparametric methods for the analysis of panel count data under informative observation and censoring times. Such methods are also applicable to a proportional rates model with cluster-specific baseline rates. However, when the censoring time and the recurrent event times are conditional independent given covariates, their methods require modeling the censoring times under the proportional hazards assumption. Most of the existing methods dealing with interval-grouped recurrent event data cannot be easily carried out using standard software. Cook and Lawless (2007) described a Poisson model with piecewise-constant rate functions for recurrent event data and illustrated the use of Poisson log-linear regression software for parameter estimation. This method assumes independent counting process increments given the covariates and, similar to Lawless and Zhan (1998), is not applicable to cluster-specific baseline rate function settings with relatively large number of clusters.

Another characteristic of the hospitalization data, common to many other recurrent event data settings, is the presence of a terminal event, i.e., an event that stops the recurrent event process (e.g., death). Models for the rate function of recurrent event data in the presence of a terminal event can generally be categorized as (1) marginal models (e.g., Ghosh and Lin, 2002; Schaubel and Zhang, 2010), which can be interpreted as the occurrence rate averaging over mortality experience, or (2) partial marginal models (e.g., Cook and Lawless, 1997; Ye, Kalbfleisch, and Schaubel, 2007), which consider the rate of the recurrent events among survivors. In this article, we consider a partial marginal model for the rate function of the recurrent event with unspecified dependence structure between the recurrent events and the terminal event.

The remainder of this article is organized as follows. In Section 2, we first propose a proportional rates model with piecewise-constant baseline rate function for clustered recurrent event data, in the absence of a terminal event. The dependence structure for within-patient events are left completely unspecified. The extension to the setting with a terminating event then follows, under a partial-marginal model. The essential parts of the estimation procedure are quite similar, although the interpretation of the covariate effects is different. The proposed estimation procedure requires only the interval-specific event and person–time totals, instead of the exact recurrent event times, which leads to considerable data reduction and hence reduced computing time. In Section 3, we compare the proposed estimation method to a joint estimating equation method (JM) based on pseudo likelihood. We derive the large-sample properties of the proposed estimators in Section 4 and assess their finite-sample performance in Section 5 under various data configurations, including settings in which the model is misspecified. In Section 6, we apply the proposed the model to the study of days hospitalized among U.S. dialysis patients. The article then concludes with some discussion in Section 7.

## 2. Model Specification and Estimation

### 2.1 *In the Absence of a Terminal Event*

As the name implies, the proposed model assumes that the baseline rate is constant over prespecified intervals and is applied to recurrent event data in the absence of terminal event. Denote the largest observation time by $\tau$. Let $a_0 < a_1 < \ldots < a_L$ denote the cut points for the $L$ intervals on $[0, \tau]$, where $a_0 = 0$, $a_L = \tau$ and $\Omega_\ell = (a_{\ell-1},\ a_\ell]$ for $\ell = 1, \ldots, L$. Let $k$ index cluster, with cluster sizes $n_1, \ldots, n_K$ and let $i$ index the subject $(i = 1, \ldots, n)$ with $n = \sum_{k=1}^{K} n_k$. For subject $i$, let $G_i$ denote cluster and let $C_i$ denote the right censoring time. Because data are often left truncated, we explicitly allow for left truncation in the formulation of the proposed methods, with left-truncation time represented by $B_i$. We then define the at-risk process by $\widetilde{Y}_i(t) = I(B_i \le t \le C_i)$ with $I(\cdot)$ being the indicator function. Let $\widetilde{N}_i^*(t)$ denote the cumulative number of events up to time $t$ and let $\widetilde{N}_i(t) = \int_0^t \widetilde{Y}_i(s) d\widetilde{N}_i^*(s)$ denote the observed number of events. We then specify the rate function for subject $i$ from cluster $k$ as

$$E\{d\widetilde{N}_i^*(t) \mid \boldsymbol{Z}_i(t), G_i = k\} = \rho_{k\ell} e^{\boldsymbol{\gamma}_0^T \boldsymbol{Z}_{i\ell}} dt,$$

where, for $\ell = 1, \ldots, L$, $\rho_{k\ell}$ is the baseline rate function for the $k$th cluster, $\boldsymbol{\gamma}_0$ is a $p$-vector parameter, $\boldsymbol{Z}_{i\ell} = \boldsymbol{Z}_i(t)$, $t \in \Omega_\ell$ is a $p$-vector possibly time-varying covariates for subject $i$. Note that any time-dependent covariates are assumed to be external (Kalbfleisch and Prentice, 2002) and constant within each interval $\Omega_\ell$. Define $G_{ik} = I(G_i = k)$, $\widetilde{Y}_{ik}(t) = G_{ik}\widetilde{Y}_i(t)$, $d\widetilde{N}_{ik}^*(t) = G_{ik} d\widetilde{N}_i^*(t)$ and $d\widetilde{N}_{ik}(t) = G_{ik} d\widetilde{N}_i(t)$. Under the assumption of independent left truncation and censoring, which can be specified as

$$\begin{aligned} &E\{d\widetilde{N}_{ik}^*(t) \mid \boldsymbol{Z}_i(t), G_i = k, \widetilde{Y}_i(t) = 1\} \\ &\quad = E\{d\widetilde{N}_{ik}^*(t) \mid \boldsymbol{Z}_i(t), G_i = k\}, \end{aligned}$$

we have

$$E\{d\widetilde{N}_{ik}(t) \mid \boldsymbol{Z}_i(t), \widetilde{Y}_{ik}(t)\} = \widetilde{Y}_{ik}(t) \rho_{k\ell} e^{\boldsymbol{\gamma}_0^T \boldsymbol{Z}_{i\ell}} dt. \tag{1}$$

### 2.2 *Piecewise-Constant Baseline Rates Model in the Presence of a Terminating Event*

When the recurrent event is potentially stopped by a terminal event (e.g., death), we can similarly specify a partial marginal model with piecewise-constant baseline rates. Let $D_i$ denote the death time for subject $i$. Define the follow-up time $X_i = C_i \wedge D_i$, with $a \wedge b = \min(a, b)$ and the at-risk process $Y_i(t) = I(B_i \leqslant t \leqslant X_i)$. Then the counting process for the recurrent events $N_i^*(t) = N_i^*(t \wedge D_i)$, which acknowledges the fact that death stops the further occurrence of recurrent events, such that $N_i^*(t)$ is a constant after $D_i$. Similar to the model in the absence of terminal event, the occurrence rate function for subject $i$ from cluster $k$ conditional on being alive is given as

$$E\{dN_i^*(t) \mid \boldsymbol{Z}_i(t), D_i \geq t, G_i = k\} = \alpha_{k\ell} e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{i\ell}} dt,$$

where, for $\ell = 1, \ldots, L$, $\alpha_{k\ell}$ is the baseline rate function for the $k$th cluster and $\boldsymbol{\beta}_0$ is a $p$-vector parameter. Define $Y_{ik}(t) = G_{ik} Y_i(t)$, $dN_{ik}^*(t) = G_{ik} dN_i^*(t)$ and $dN_{ik}(t) = G_{ik} dN_i(t)$. Under the assumption of independent left truncation and censoring, which is written as

$$\begin{aligned} E\{dN_{ik}^*(t) \mid \boldsymbol{Z}_i(t), Y_i(t) = 1, G_i = k\} \\ = E\{dN_{ik}^*(t) \mid \boldsymbol{Z}_i(t), D_i \geq t, G_i = k\}, \end{aligned}$$

we have

$$E\{dN_{ik}(t) \mid \boldsymbol{Z}_i(t), Y_{ik}(t)\} = Y_{ik}(t) \alpha_{k\ell} e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{i\ell}} dt. \quad (2)$$

### 2.3 *Estimation*

Next, we describe the estimation method for the model in the presence of a terminal event. Similar estimating procedure can be applied to the model in the absence of terminating event by setting $D_i = \tau$. We first define some notation. For subject $i$ (from cluster $k$), let $t_{ik\ell} = \int_{a\ell-1}^{a\ell} Y_{ik}(t) dt$ denote the time at risk (exposure time) and $d_{ik\ell} = \int_{a\ell-1}^{a\ell} dN_{ik}(t)$ be the observed number of events experienced in $\Omega_\ell$. In addition, for $r = 0, 1, 2$, $k = 1, \ldots, K$ and $\ell = 1, \ldots, L$, we define

$$\begin{aligned} \boldsymbol{S}_{k\ell}^{(r)}(\boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_{i\ell}^{\otimes r} t_{ik\ell} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{i\ell}}, \\ \overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta}) &= \boldsymbol{S}_{k\ell}^{(1)}(\boldsymbol{\beta}) / S_{k\ell}^{(0)}(\boldsymbol{\beta}), \\ \boldsymbol{V}_{k\ell}(\boldsymbol{\beta}) &= \boldsymbol{S}_{k\ell}^{(2)}(\boldsymbol{\beta}) / S_{k\ell}^{(0)}(\boldsymbol{\beta}) - \overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta})^{\otimes 2}, \end{aligned}$$

where $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. We next define the compensated counting process,

$$dM_{ik}(t) = dN_{ik}(t) - Y_{ik}(t) \alpha_{k\ell} e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{i\ell}} dt, \quad t \in \Omega_\ell,$$

for $\ell = 1, \ldots, L$. By the specification of the model (2) and under the corresponding independent left truncation and censoring assumption, $E\{dM_{ik}(t) \mid \boldsymbol{Z}_{i\ell}, Y_{ik}(t)\} = 0$ for $t \in \Omega_\ell$. Thus, it follows that

$$\xi_{ik\ell}(\boldsymbol{\beta}_0) = \int_{a_{\ell-1}}^{a_\ell} dM_{ik}(t) = d_{ik\ell} - \alpha_{k\ell} t_{ik\ell} e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{i\ell}},$$

has mean zero because

$$E\left\{\int_{a_{\ell-1}}^{a_\ell} dM_{ik}(t)\right\} = E\left[\int_{a_{\ell-1}}^{a_\ell} E\left\{dM_{ik}(t) \mid \boldsymbol{Z}_{i\ell}, Y_{ik}(t)\right\}\right] = 0.$$

We consider the estimating function,

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \{\boldsymbol{Z}_{i\ell} - \overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta})\} \xi_{ik\ell}(\boldsymbol{\beta}),$$

motivated by the fact that $\boldsymbol{U}(\boldsymbol{\beta}_0)$ can be shown to have mean 0 asymptotically, which can be proved by replacing $\overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta})$ with the corresponding limiting values in $\boldsymbol{U}(\boldsymbol{\beta})$. We can simplify $\boldsymbol{U}(\boldsymbol{\beta})$ to

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{\ell=1}^{L} \sum_{k=1}^{K} \{\boldsymbol{Z}_{i\ell} - \overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta})\} d_{ik\ell}, \quad (3)$$

such that an estimator for $\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}$, can be obtained by solving $\boldsymbol{U}(\boldsymbol{\beta}) = 0$. A Breslow–Aalen type estimator for $\alpha_{k\ell}$ is then given as

$$\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}}) = \frac{d_{\bullet k\ell}}{n S_{k\ell}^{(0)}(\widehat{\boldsymbol{\beta}})}, \quad (4)$$

where $\bullet$ denotes the summation over the corresponding subscript. Therefore, the corresponding Breslow–Aalen type estimator for the cumulative baseline rate function $\mu_{0k}(t) = \sum_{\ell=1}^{L} \alpha_{k\ell}(a_\ell \wedge t - a_{\ell-1} \wedge t)$ is given as

$$\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}) = \sum_{\ell=1}^{L} \frac{d_{\bullet k\ell}}{n S_{k\ell}^{(0)}(\widehat{\boldsymbol{\beta}})} (a_\ell \wedge t - a_{\ell-1} \wedge t), \quad (5)$$

One may notice that (3) is similar to the partial score equation for recurrent event data except for an offset term, and a weight term. Therefore, the proposed estimation method is easy to implement with `SAS` (`PROC PHREG`) or `R` (`coxph`) with the censoring variable $d_{ik\ell}$, the weight term $w_{ik\ell} = \max(d_{ik\ell}, 1)$, and the offset term $\log(t_{ik\ell}) - \log(w_{ik\ell})$. It should also be noted that, unlike the conventional partial score equation in which statistics are computed at each distinct recurrent event time, the proposed estimating equation is calculated only for each interval, which greatly speeds up the calculation, especially when the number of event occurrences is large.

A few additional notes are in order. First, although the models for the recurrent event data are different in the absence and the presence of a terminal event, the estimation methods are essentially the same upon redefinition of the at-risk indicators. The main difference lies in the interpretation of the covariate effect. Second, if the data are not left truncated, the proposed methods can be applied by setting $B_i = 0$ for all $i = 1, \ldots, n$. Third, in the absence of terminating event the unbiasedness of $\boldsymbol{U}(\boldsymbol{\beta}_0)$ can be proved based on conditional expectation arguments (e.g., Appendix 7.1 in Schaubel and Cai, 2005). Finally, we emphasize cluster-specific baseline rates model in this article. When the baseline rate function is common to all clusters, an analogous estimation procedure can be carried out with $\boldsymbol{S}_{k\ell}^{(r)}$ and $d_{\bullet k\ell}$ ($k = 1, \ldots, K$, $\ell = 1, \ldots, L$) replaced by the corresponding quantities summing over all the clusters in (3) and (4).

## 3. Comparison with Joint Estimating Equation Approach

An alternative estimation approach is based on pseudo-likelihood that ignores within-subject and within-cluster dependence. Let $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{1L}, \ldots, \alpha_{K1}, \ldots, \alpha_{KL})'$ and

$\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$. The pseudo-likelihood function is then $L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{\ell=1}^{L} L_{ik\ell}(\boldsymbol{\theta})$, where $L_{ik\ell}(\boldsymbol{\theta})$ is given as $L_{ik\ell}(\boldsymbol{\theta}) = (\alpha_{k\ell}e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{i\ell}})^{d_{ik\ell}} e^{-\alpha_{k\ell}t_{ik\ell}e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{i\ell}}}$. The resulting log likelihood is then

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left\{d_{ik\ell}(\log\alpha_{k\ell} + \boldsymbol{\beta}^T\boldsymbol{Z}_{i\ell}) - \alpha_{k\ell}t_{ik\ell}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{i\ell}}\right\}.$$

The score vector $\boldsymbol{U}^J(\boldsymbol{\theta}) = \{\boldsymbol{U}_{\boldsymbol{\alpha}}^T(\boldsymbol{\theta}), \boldsymbol{U}_{\boldsymbol{\beta}}^T(\boldsymbol{\theta})\}^T$, with $\boldsymbol{U}_{\boldsymbol{\alpha}} = (U_{\alpha_{11}}, \ldots, U_{\alpha_{KL}})^T$, can be obtained by taking the partial derivative of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ as

$$U_{\alpha_{k\ell}}(\boldsymbol{\theta}) = \frac{d_{\bullet k\ell}}{\alpha_{k\ell}} - nS_{k\ell}^{(0)}(\boldsymbol{\beta}), \quad k = 1, \ldots, K, \ell = 1, \ldots, L, \quad (6)$$

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\theta) = \sum_{i=1}^{n}\sum_{\ell=1}^{L}\sum_{k=1}^{K}\boldsymbol{Z}_{i\ell}\xi_{ik\ell}(\boldsymbol{\beta}). \quad (7)$$

The solution of the joint estimating equation $\boldsymbol{U}^J(\theta) = 0$, $\widehat{\boldsymbol{\theta}}$, is then an estimator for $\boldsymbol{\theta}$. It can be easily seen that this JM gives the same estimator as the proposed method (PM). A profile estimator $\widetilde{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ for $\boldsymbol{\alpha}$ can be obtained from (6) given $\boldsymbol{\beta}$, which equals the Breslow–Aalen estimator from PM. Replacing $\boldsymbol{\alpha}$ with $\widetilde{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ in (7) then gives the same estimating function (3) for $\boldsymbol{\beta}$ in PM. Moreover, unlike PM that calculates the estimated covariance matrix for $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ separately, JM estimates the joint covariance matrix for $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, which involves inverting the observed information matrix $\boldsymbol{I}^J$. As the minus second partial derivative of $\ell(\boldsymbol{\theta})$, $\boldsymbol{I}^J$ is of dimension ($KL$ + $p$) with the upper left square submatrix corresponding to $\alpha$ being a diagonal matrix with the $\{k(L-1) + \ell\}$th diagonal element equal to $d_{\bullet k\ell}/\alpha_{k\ell}^2$. When $d_{\bullet k\ell} = 0$, which is quite possible for clusters with small cluster size or less frequent recurrent events in interval $\Omega_\ell$, $I^J$ is not positive definite. As a result, JM cannot give an estimator for the joint covariance matrix.

## 4. Asymptotic Properties

The asymptotic properties are derived for the model in the presence of a terminal event. As illustrated in Section 2.3, in the absence of terminal event, one can obtain similar results by letting $D_i = \tau$.

For $i = 1, \ldots, n$, we impose the following regularity conditions:

(a) $\{N_i(t), Y_i(t), \boldsymbol{Z}_i(t), G_i\}_{i=1}^n$ are independent and identically distributed;
(b) $P\{G_{ik} = 1\} \in (0, 1]$.
(c) $E\{Y_i(t)\} > 0$, for all $t \in (0, \tau]$.
(d) $N_i(t)$, are bounded by a constant.
(e) $\boldsymbol{Z}_{i\ell}$, $\ell = 1, \ldots, L$ are bounded by a constant.
(f) Let $\mathcal{B}$ be a neighborhood of $\boldsymbol{\beta}_0$. For $d = 0, 1, 2$, $\boldsymbol{s}_{k\ell}^{(d)}(\boldsymbol{\beta})$ are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$, where $\boldsymbol{s}_{k\ell}^{(d)}(\boldsymbol{\beta})$ is the limiting values of $\boldsymbol{S}_{k\ell}^{(d)}(\boldsymbol{\beta})$; $\boldsymbol{s}_{k\ell}^{(1)}(\boldsymbol{\beta})$ and $\boldsymbol{s}_{k\ell}^{(2)}(\boldsymbol{\beta})$ are bounded and $s_{k\ell}^{(0)}(\boldsymbol{\beta})$ is bounded away from 0 on $\mathcal{B}$ with

$$\boldsymbol{s}_{k\ell}^{(1)}(\boldsymbol{\beta}) = \frac{\partial}{\partial\boldsymbol{\beta}}s_{k\ell}^{(0)}(\boldsymbol{\beta}),\ \boldsymbol{s}_{k\ell}^{(2)}(\boldsymbol{\beta}) = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}s_{k\ell}^{(0)}(\boldsymbol{\beta}).$$

(g) Positive definiteness of the matrix

$$\boldsymbol{A} = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\alpha_{k\ell}\boldsymbol{v}_{k\ell}(\boldsymbol{\beta}_0)s_{k\ell}^{(0)}(\boldsymbol{\beta}_0),$$

where $\boldsymbol{v}_{k\ell}(\boldsymbol{\beta}) = \boldsymbol{s}_{k\ell}^{(2)}(\boldsymbol{\beta})/s_{k\ell}^{(0)}(\boldsymbol{\beta}) - \bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta})^{\otimes 2}$ and $\bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta}) = \boldsymbol{s}_{k\ell}^{(1)}(\boldsymbol{\beta})/s_{k\ell}^{(0)}(\boldsymbol{\beta})$.

Assumption (a) specifies that the independent units in the proposed method are subjects. Assumption (b) states that the probability of a randomly selected subject being assigned to a cluster is nonzero for any cluster. Both conditions are necessary so that parameter estimators for the cluster-specific baseline rate functions are estimable for all clusters.

We next summarize the theoretical results for $\widehat{\boldsymbol{\beta}}$ by the following theorem.

Theorem 1. *Under regularity conditions* $(a) - (g), \widehat{\boldsymbol{\beta}}$ *converges almost surely to* $\boldsymbol{\beta}_0$ *as* $n \to \infty$, *while* $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *converges to ap-variate normal vector with mean 0 and covariance matrix* $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{\Xi}\boldsymbol{A}^{-1}$, *where* $\boldsymbol{\Xi} = E\{\boldsymbol{U}_1(\boldsymbol{\beta}_0)^{\otimes 2}\}$, *with*

$$\boldsymbol{U}_i(\boldsymbol{\beta}) = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\{\boldsymbol{Z}_{i\ell} - \bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta})\}\xi_{ik\ell}(\boldsymbol{\beta}).$$

A consistent estimator for $\boldsymbol{\Sigma}$ can be obtained by replacing $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ with their empirical counterparts.

Theorem (1) can be proved by combining the uniform strong law of large numbers and the central limit theorem, as is shown in the Appendix. We next present the essential asymptotic properties for $\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}})$.

Theorem 2. *Under regularity conditions* $(a) - (f)$, *for* $k = 1, \ldots, K$, $\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}})$ *converges almost surely to* $\mu_{0k}(t)$ *uniformly in* $t \in [0, \tau]$; *the processn*$^{1/2}\{\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}) - \mu_{0k}(t)\}$ *converges to a zero-mean Gaussian process with covariance function* $\Psi_k(s, t) = E\{\psi_{1k}(s)\psi_{1k}(t)\}$, *where*

$$\psi_{ik}(t) = \sum_{\ell=1}^{L}\psi_{ik\ell}(\boldsymbol{\beta}_0)(a_\ell \wedge t - a_{\ell-1} \wedge t),$$

$$\psi_{ik\ell}(\boldsymbol{\beta}) = \frac{\xi_{ik\ell}(\boldsymbol{\beta})}{s_{k\ell}^{(0)}(\boldsymbol{\beta})} - \alpha_{k\ell}\bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta})\boldsymbol{A}^{-1}\boldsymbol{U}_i(\boldsymbol{\beta}_0).$$

We show in Appendix that $n^{1/2}\{\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}) - \mu_{0k}(t)\}$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\psi_{ik}(t)$. A consistent estimator for $\Psi_k(s, t)$ is then $n^{-1}\sum_{i=1}^{n}\widehat{\psi}_{ik}(s)\widehat{\psi}_{ik}(t)$, where $\widehat{\psi}_{ik}(t)$ is the empirical measure of $\psi_{ik}(t)$.

## 5. Simulation Study

In this section, we conduct simulation studies to assess the finite sample performance of the proposed estimators. First, we consider $L = 3, 6, 12$ intervals for the baseline rate function with one set of prespecified cut points. Lawless and Zhan (1998) suggested that 'In most practical situations, it is satisfactory to use piecewise-constant intensities with 4–10 pieces.' In this stage, we aim to verify this statement by comparing the proposed estimators using different numbers of intervals under both the true and a misspecified model. Next, we choose another set of cut points and redo the simulation in the first stage. By comparing the results between the two stages, we seek appropriate guidelines for the location of cut points.

### 5.1 *Finite Sample Performance of the Estimators with Different Number of Intervals*

Simulation studies were conducted to assess the performance of the estimation method in the presence of a terminal event. The same settings could be used to evaluate the finite sample performance of the estimators in the absence of terminal event by defining a new censoring time $\widetilde{C} = \min(D, C)$.

In the first simulation study, for subject $(i, k)$, we generate recurrent event times from

$$E\{dN_{ik}^{*}(t)\,|\,\boldsymbol{Z}_i, W_i, D_i \geq t\} = W_i e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i} d\mu_{0k}(t),$$

where the subject-level random effect $W_i$ follows gamma distribution with unit mean and variance $\sigma^2 = 1$, and the cluster-specific cumulative baseline rate function $\mu_{0k}(t) = t$, $\boldsymbol{Z}_i = \{Z_{i1}, Z_{i2}\}$ are two-vector covariates with $Z_{i1} \sim$ Bernoulli(0.5) and $Z_{i2} \sim N(0, 0.25)$, $\beta_1 = 0.5$ and $\beta_2 = 0.25, 0.5, 0.75, 1$. In addition, we let $D_i \sim \text{Exp}(0.1 + 0.1Z_{i1})$ and $C_i \sim U(5, 10)$. The average number of recurrent events ranged from six to eight. We set $K = 50, 100$ and $n_k = 20, 50, 100$. For each simulated data set, we estimated $\boldsymbol{\beta}_0$ under model (2) with three settings for the piecewise-constant baseline rate function: the first setting is with $L = 3$ pieces defined by 0, 2, 4, 10; the second setting is with $L = 6$ pieces defined by $0, 1, \ldots, 5, 10$; the third setting is with $L = 12$ pieces resulting from adding six midpoints of the intervals in the second setting. The results are shown in Table 1 for $\widehat{\beta}_1$ based on 1000 simulations.

For the first simulation study where the true model is actually piecewise constant, we do not present the results based on $L = 3$ because they are similar to results for $L = 6$ and $L = 12$. For all of the data configurations in Table 1, the estimator for $\beta_1$ corresponding to the binary covariate, $\widehat{\beta}_1$, is approximately unbiased with the bias reduced with increasing cluster size. The mean of the asymptotic standard error (ASE) of $\widehat{\beta}_1$ is generally close to the empirical standard deviation (ESD) of $\widehat{\beta}_1$, and the coverage probabilities (CP) are fairly close to the nominal value. Adding more cut points does not seem to improve the performance of the estimator. Results for $\widehat{\beta}_2$ are similar to $\widehat{\beta}_1$ and thus are not provided. By comparing the results with different combinations of $K$ and $n_k$ , we also notice that the performance of the estimators with the same total sample size ($K = 50$, $n_k = 100$; $K = 100$, $n_k = 50$) are similar regardless of the specific configuration of the number of centers and center size.

In the second simulation study, we let $\mu_{0k}(t) = 0.5t^2$ and left other settings unchanged. The average number of recurrent events per subject ranged from 18 to 20. For each setting, 1000 data sets are simulated. The results for $\widehat{\beta}_1$ are shown in Table 2, again based on 1000 replicates.

Under misspecification of the baseline rate function, the estimator $\widehat{\beta}_1$ is biased based on the percentage of bias (% BIAS) with the bias reduced by adding more pieces to the baseline rate function. Although the mean square error (MSE) does not seem to be improved, the CP gets closer

**Table 1**
*Results of* $\widehat{\boldsymbol{\beta}}_1$ *in the first simulation study with* $\beta_1 = 0.5$, $\mu_{0k}(t) = t$, *and* 1000 *replicates*

| | | | $L = 6$ | | | | $L = 12$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $n_k$ | $\beta_2$ | BIAS | ASE | ESD | CP | BIAS | ASE | ESD | CP |
| 50 | 20 | 1 | −0.005 | 0.081 | 0.082 | 95.2 | 0.005 | 0.081 | 0.082 | 95.2 |
| | | 0.75 | 0.004 | 0.079 | 0.079 | 94.9 | 0.005 | 0.079 | 0.079 | 94.9 |
| | | 0.5 | 0.004 | 0.078 | 0.078 | 94.5 | 0.005 | 0.077 | 0.078 | 94.2 |
| | | 0.25 | 0.003 | 0.077 | 0.075 | 95.0 | 0.004 | 0.077 | 0.075 | 94.8 |
| 50 | 50 | 1 | 0.006 | 0.053 | 0.051 | 96.4 | 0.006 | 0.053 | 0.051 | 96.2 |
| | | 0.75 | 0.004 | 0.052 | 0.050 | 96.1 | 0.004 | 0.052 | 0.050 | 96.2 |
| | | 0.5 | 0.004 | 0.051 | 0.049 | 95.9 | 0.004 | 0.050 | 0.049 | 95.9 |
| | | 0.25 | 0.010 | 0.045 | 0.042 | 96.2 | 0.010 | 0.045 | 0.042 | 96.0 |
| 50 | 100 | 1 | 0.010 | 0.037 | 0.037 | 94.9 | 0.010 | 0.037 | 0.037 | 94.8 |
| | | 0.75 | 0.009 | 0.036 | 0.035 | 95.2 | 0.010 | 0.036 | 0.035 | 95.0 |
| | | 0.5 | 0.009 | 0.035 | 0.033 | 95.3 | 0.009 | 0.035 | 0.033 | 95.4 |
| | | 0.25 | 0.010 | 0.034 | 0.033 | 94.4 | 0.010 | 0.034 | 0.033 | 94.6 |
| 100 | 20 | 1 | −0.009 | 0.058 | 0.056 | 95.6 | −0.009 | 0.058 | 0.056 | 95.5 |
| | | 0.75 | −0.010 | 0.056 | 0.056 | 96.0 | −0.010 | 0.056 | 0.056 | 95.6 |
| | | 0.5 | −0.010 | 0.055 | 0.055 | 95.0 | −0.010 | 0.055 | 0.054 | 95.1 |
| | | 0.25 | −0.010 | 0.055 | 0.053 | 95.4 | −0.010 | 0.055 | 0.053 | 95.6 |
| 100 | 50 | 1 | 0.009 | 0.038 | 0.038 | 95.3 | 0.009 | 0.038 | 0.038 | 95.4 |
| | | 0.75 | 0.008 | 0.035 | 0.037 | 94.9 | 0.009 | 0.037 | 0.037 | 94.7 |
| | | 0.5 | 0.008 | 0.036 | 0.035 | 94.8 | 0.008 | 0.036 | 0.034 | 95.0 |
| | | 0.25 | 0.009 | 0.035 | 0.033 | 95.2 | 0.009 | 0.035 | 0.034 | 95.3 |
| 100 | 100 | 1 | −0.002 | 0.027 | 0.027 | 95.8 | 0.002 | 0.027 | 0.027 | 95.6 |
| | | 0.75 | 0.001 | 0.026 | 0.025 | 96.4 | 0.001 | 0.026 | 0.025 | 96.1 |
| | | 0.5 | 0.001 | 0.026 | 0.024 | 96.0 | 0.001 | 0.026 | 0.024 | 96.0 |
| | | 0.25 | 0.002 | 0.025 | 0.024 | 96.4 | 0.002 | 0.025 | 0.024 | 96.4 |

**Table 2**
*Results of the second simulation study for $\widehat{\beta}_1$ with $\beta_1 = 0.5$, $\mu_{0k}(t) = 0.5t^2$, and 1000 replicates*

| | | | $L = 3$ | | | $L = 6$ | | | $L = 12$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $n_k$ | $\beta_2$ | % BIAS | MSE | CP | % BIAS | MSE | CP | % BIAS | MSE | CP |
| 50 | 20 | 1 | −7.7 | 0.125 | 91.4 | −4.5 | 0.130 | 92.6 | −2.3 | 0.129 | 93.1 |
| | | 0.75 | −7.6 | 0.125 | 91.4 | −4.3 | 0.125 | 92.9 | −2.2 | 0.124 | 94.4 |
| | | 0.5 | −7.5 | 0.122 | 91.6 | −4.2 | 0.122 | 93.1 | −2.1 | 0.121 | 92.9 |
| | | 0.25 | −7.7 | 0.121 | 91.8 | −4.4 | 0.120 | 93.4 | −2.3 | 0.119 | 94.6 |
| 50 | 50 | 1 | −5.8 | 0.084 | 92.9 | −2.8 | 0.084 | 95.6 | −1.2 | 0.083 | 96.0 |
| | | 0.75 | −5.9 | 0.081 | 93.4 | −2.9 | 0.081 | 96.0 | −1.2 | 0.080 | 96.3 |
| | | 0.5 | −5.7 | 0.079 | 92.8 | −2.7 | 0.079 | 95.9 | −1.1 | 0.078 | 96.2 |
| | | 0.25 | −5.9 | 0.077 | 92.4 | −2.9 | 0.077 | 95.8 | −1.2 | 0.077 | 95.1 |
| 50 | 100 | 1 | −5.1 | 0.061 | 90.6 | −2.1 | 0.061 | 95.0 | −0.5 | 0.061 | 96.2 |
| | | 0.75 | −5.0 | 0.058 | 91.6 | −2.0 | 0.058 | 95.2 | −0.4 | 0.058 | 96.3 |
| | | 0.5 | −5.1 | 0.056 | 90.6 | −2.0 | 0.056 | 94.6 | −0.4 | 0.056 | 95.7 |
| | | 0.25 | −5.1 | 0.055 | 91.0 | −2.0 | 0.055 | 95.7 | −0.4 | 0.055 | 96.6 |
| 100 | 20 | 1 | −9.7 | 0.092 | 87.0 | −6.7 | 0.091 | 90.9 | −5.0 | 0.090 | 93.3 |
| | | 0.75 | −9.7 | 0.089 | 86.8 | −6.6 | 0.088 | 91.1 | −5.0 | 0.088 | 93.0 |
| | | 0.5 | −9.6 | 0.087 | 87.4 | −6.5 | 0.086 | 91.5 | −4.9 | 0.086 | 93.4 |
| | | 0.25 | −9.8 | 0.086 | 87.6 | −6.7 | 0.086 | 91.3 | −5.0 | 0.085 | 93.0 |
| 100 | 50 | 1 | −5.3 | 0.061 | 88.9 | −2.3 | 0.060 | 94.4 | −0.6 | 0.060 | 95.5 |
| | | 0.75 | −5.2 | 0.058 | 89.9 | −2.2 | 0.058 | 94.1 | −0.5 | 0.058 | 94.5 |
| | | 0.5 | −5.3 | 0.056 | 90.2 | −2.2 | 0.056 | 93.9 | −0.6 | 0.056 | 94.3 |
| | | 0.25 | −5.3 | 0.055 | 89.9 | −2.2 | 0.055 | 95.0 | −0.6 | 0.054 | 95.6 |
| 100 | 100 | 1 | −5.3 | 0.044 | 86.4 | −2.4 | 0.044 | 93.1 | −0.6 | 0.043 | 94.9 |
| | | 0.75 | −5.3 | 0.041 | 86.9 | −2.4 | 0.041 | 93.7 | −0.5 | 0.041 | 95.4 |
| | | 0.5 | −5.3 | 0.040 | 85.8 | −2.4 | 0.040 | 93.6 | −0.5 | 0.040 | 96.0 |
| | | 0.25 | −5.3 | 0.039 | 85.0 | −2.4 | 0.039 | 93.9 | −0.5 | 0.039 | 96.0 |

to the nominal level as the number of pieces in the assumed baseline rate function increases. The improvement is more obvious comparing $L = 3$ and $L = 12$. Results for $\widehat{\beta}_2$ are very similar to those for $\widehat{\beta}_1$ and, hence, are not presented.

Simulation studies with a variety of center sizes (including one that resembles the motivating example) are also considered for the evaluation of the regression parameter estimators. For example, we conduct numerical studies with $K = 50$ and center sizes of 5, 10, 20, 50, and 100 each for 10 centers. The results are similar to those for the equal-center-size settings and are presented in Table S.1 of the Supplementary Web Materials.

In the third simulation study, we assess the asymptotic properties for $\widehat{\mu}_{0k}(t)$ with $\boldsymbol{\beta}_0 = (0.5, 1)$ and $K = 50$ under two scenarios; one in which $\mu_{0k}(t) = t$ and a second in which $\mu_{0k}(t) = 0.5\, t^2$. Remaining characteristics are as described previously. We let $n_1 = 20$, $n_2 = 50$, $n_3 = 100$, and $n_k = 50$ for $k = 4, \ldots, 50$. In both cases, the rates are assumed to be piecewise constant. Specifically, we estimated $\mu_{0k}(t)$ under model (2) with two settings for the cut points of the piecewise-constant baseline rate function: (i) with $L = 6$ and $a_\ell = 0, 1, \ldots, 5, 10$ (ii) with $L = 12$ and $a_\ell = 0, 0.5, 1, 1.5, \ldots, 5, 7.5, 10$. Thus, in the second setting, we double the number of cut points by including the midpoints of all the intervals from the first setting. We then evaluate the performance of the estimator for $\mu(t)$ under each scenario at four selected time points 1.51, 3.56, 5.56, 7.23, which result in 80%, 60%, 40%, and 20% of the subjects at risk, respectively. For each setting, 1000 data sets are simulated.

Results for the third simulation study are shown in Table 3. From Table 3a, the estimator for $\mu(t)$ is approximately unbiased under the piecewise-constant baseline rates model. The ASE is generally similar to the ESD, and the CP is close to 95%. In addition, the piecewise-constant baseline rates model with $L = 12$ does not seem to produce a better estimator than the model with $L = 6$ for $\mu_{0k}(t)$ at the four selected time points, in terms of unbiasedness and efficiency. The performance of the estimators get worse as time increases. The results from the linear baseline rates model are presented in Table 3b. In general, there is some bias in $\widehat{\mu}_{0k}(t)$ and the % BIAS increases with time, which is not surprising because the expected number of subjects at risk $R_k(t) = n_k E\{Y(t)\}$ also decreases with time. In addition, the bias is reduced by adding more pieces in the assumed baseline rate function, although the MSE and CP do not seem to be improved. Overall, the bias is reduced and the efficiency is improved with increasing center size.

In summary, the proposed method seems to work well for the proportional rates model even when the underling baseline rate function is not piecewise constant. In addition, using $L = 3$ pieces does not seem to provide a good estimates for the regression parameters. The performance of the estimators under the model with six pieces for the baseline rate function

**Table 3**

*Results of $\widehat{\mu}_{0k}(t)$ in the third simulation study with $\beta_1 = 0.5$, $\beta_2 = 1$, and 1000 replicates*

(a) $\mu_{0k}(t) = t$

| | | | | | $L = 6$ | | | | $L = 12$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $n_k$ | $t$ | $R_k(t)$ | $\mu_{0k}(t)$ | BIAS | ASE | ESD | CP | BIAS | ASE | ESD | CP |
| 1 | 20 | 1.51 | 16 | 1.51 | 0.017 | 0.414 | 0.485 | 88.9 | 0.016 | 0.424 | 0.493 | 89.9 |
| | | 3.56 | 12 | 3.56 | 0.042 | 0.939 | 1.112 | 89.0 | 0.045 | 0.945 | 1.125 | 89.1 |
| | | 5.56 | 8 | 5.56 | 0.037 | 1.479 | 1.820 | 86.4 | 0.039 | 1.483 | 1.816 | 86.3 |
| | | 7.23 | 4 | 7.23 | 0.034 | 1.953 | 2.556 | 84.0 | 0.034 | 1.972 | 2.528 | 84.8 |
| 2 | 50 | 1.51 | 40 | 1.51 | −0.004 | 0.272 | 0.297 | 91.0 | −0.003 | 0.278 | 0.305 | 91.7 |
| | | 3.56 | 30 | 3.56 | 0.026 | 0.623 | 0.694 | 91.2 | 0.025 | 0.627 | 0.701 | 91.3 |
| | | 5.56 | 20 | 5.56 | 0.032 | 1.011 | 1.127 | 91.6 | 0.033 | 1.014 | 1.131 | 91.9 |
| | | 7.23 | 10 | 7.23 | 0.025 | 1.373 | 1.514 | 91.3 | 0.039 | 1.381 | 1.522 | 91.6 |
| 3 | 100 | 1.51 | 80 | 1.51 | 0.004 | 0.198 | 0.208 | 93.9 | 0.003 | 0.203 | 0.209 | 93.9 |
| | | 3.56 | 60 | 3.56 | 0.001 | 0.454 | 0.483 | 93.4 | −0.001 | 0.457 | 0.486 | 93.7 |
| | | 5.56 | 40 | 5.56 | 0.020 | 0.735 | 0.792 | 92.6 | 0.024 | 0.736 | 0.791 | 92.6 |
| | | 7.23 | 20 | 7.23 | 0.020 | 0.996 | 1.099 | 91.7 | 0.024 | 0.997 | 1.089 | 92.1 |

(b) $\mu_{0k}(t) = 0.5t^2$

| | | | | | $L = 6$ | | | $L = 12$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $n_k$ | $t$ | $R_k(t)$ | $\mu_{0k}(t)$ | % BIAS | MSE | CP | % BIAS | MSE | CP |
| 1 | 20 | 1.51 | 16 | 1.14 | 5.4 | 0.539 | 89.6 | 1.8 | 0.530 | 89.9 |
| | | 3.56 | 12 | 6.16 | 3.4 | 2.647 | 89.1 | 1.4 | 2.622 | 88.5 |
| | | 5.56 | 8 | 15.46 | 4.4 | 7.126 | 85.6 | 2.9 | 6.961 | 86.0 |
| | | 7.23 | 4 | 26.11 | 3.1 | 13.035 | 82.8 | −1.2 | 12.205 | 82.6 |
| 2 | 50 | 1.51 | 40 | 1.14 | 5.3 | 0.344 | 90.5 | 0.1 | 0.329 | 92.5 |
| | | 3.56 | 30 | 6.16 | 2.9 | 1.699 | 91.8 | 1.0 | 1.675 | 91.4 |
| | | 5.56 | 20 | 15.46 | 5.4 | 4.601 | 91.8 | 3.8 | 4.524 | 92.2 |
| | | 7.23 | 10 | 26.11 | 2.6 | 8.352 | 90.2 | 0.1 | 7.880 | 90.7 |
| 3 | 100 | 1.51 | 80 | 1.14 | 4.6 | 0.247 | 90.3 | 0.7 | 0.236 | 93.1 |
| | | 3.56 | 60 | 6.16 | 3.4 | 1.211 | 92.4 | 1.5 | 1.199 | 92.5 |
| | | 5.56 | 40 | 15.46 | 5.4 | 3.338 | 91.6 | 3.9 | 3.269 | 91.4 |
| | | 7.23 | 20 | 26.11 | 1.1 | 6.185 | 91.7 | 0.2 | 5.791 | 92.0 |

seems to be fairly good. The % BIAS is generally below 5% for both the regression parameter estimators and the cumulative baseline rates estimators. Although adding more pieces does seem to reduce the bias, it does not seem to improve the MSE and the CP. In conclusion, we think Lawless and Zhan's (1998) suggestion on using 4–10 pieces is reasonable, and we suggest using at least six pieces for the % BIAS to be controlled under a reasonable threshold.

### 5.2 *Location of Cut Points*

Based on the previous simulation studies, we suggest including at least six pieces in the assumed baseline rate function. The next question is then how to decide the location of the cut points. If the trend of the recurrence rate is known based on previous literature, the location of the cut points could be specified such that the main trends in the rate function are captured. For example, if it is known that recurrent events occur more frequently in the early stage of follow-up, making finer intervals at earlier follow-up times is recommended. Alternatively, without previous knowledge of the nature of the rate function, we suggest choosing cut points based on the observed-data cumulative rate functions, which results in approximately equal number of recurrent events across intervals. Such a strategy helps ensure that there are sufficient data within each interval.

We repeated the first and the second simulation studies described in the previous subsection with a new set of cut points $a_\ell : E\{N(a_\ell)\} = \frac{\ell}{L}E\{N(\tau)\}$, $\ell = 1, \ldots, L$. Under the piecewise-constant model with $\mu_{0k}(t) = t$ as well as exponentially distributed terminating event time and uniformly distributed censoring time, the new cut points with $L = 6$ are 0.77, 1.66, 2.68, 3.91, 5.54, 10. On the other hand, when the true underlying model has linear baseline rates with $\mu_{0k}(t) = 0.5t^2$, the new set of cut points with $L = 6$ are 2.40, 3.60, 4.64, 5.65, 6.92, 10. The results of the new estimators $\widehat{\beta}_1^{\text{new}}$ and $\widehat{\mu}_{0k}^{\text{new}}(t)$ are presented in Table S.2 and Table S.3 of the Supplementary Materials. For the underlying model with piecewise-constant baseline rates, using the new cut points does improve the CPs for the estimators, although the bias and the efficiency are relatively the same as using the old cut points. Under the misspecified model with linear baseline rates, the recommended selection method greatly reduces the bias. The % BIAS with $L = 6$ using the new set of cut points is comparable to that with $L = 12$ using the old set. This is because using the recommended selection method helps to balance the data used in each interval.

### 5.3 *Computational Advantages*

Because the motivation of the proposed method is to handle recurrent event data with large event occurrence rate, it is also of interest to evaluate the computational advantage of the proposed method relative to the common semiparametric proportional rates model. We conduct a simulation study using the same setting as in the first simulation study with $\beta_1 = 0.5$, $\beta_2 = 1$ and $\mu_{0k}(t) = 0.5t,\ t,\ 1.5t,\ 2t$. The average number of observed recurrent events for each subject is 4, 8, 12, and 16, respectively. We consider $K = 200$ centers with center size of $n_k = 100$ subjects resulting a total of 20,000 subjects. In Table S.4, we summarize the computation time using the common semiparametric method (CM) and the PM on a DELL Optiplex 780 PC based on 100 replicates. The computation time using the PM is quite stable around 1.2 minutes regardless of the average number of recurrent events. On the other hand, the computation time using the CM keeps increasing from 17 minutes to about 2 hours. The PM is 13 times faster than the CM even with an average of four events per subject and the ratio for the computation time between the two methods increases to 87 with an average of 16 recurrent events per subject. In practice, it takes the PM about 2–3 hours to fit the piecewise-constant baseline rates model to the motivating example with an average of 22 days in hospital. By extension, it would take the CM several weeks to run one fit for the motivating example. Note that, although we contrast computing times using a desktop computer, it is clear that the computational issues associated with the CM would be substantial with any computer platform.

**Table 4**
*Analysis of hospitalization days for Medicare dialysis patients in the United States*

| Covariates | Estimates | SE | p-value |
|---|---|---|---|
| Age (in years) | | | |
| 0–24 | 0.057 | 0.007 | <0.0001 |
| 25–44 | −0.172 | 0.003 | <0.0001 |
| 45–59 | −0.217 | 0.001 | <0.0001 |
| 60–74 | 0 | . | . |
| 75+ | 0.093 | 0.001 | <0.0001 |
| Race | | | |
| Native | −0.089 | 0.003 | <0.0001 |
| Asian | −0.350 | 0.002 | <0.0001 |
| African-American | −0.064 | 0.001 | <0.0001 |
| Other | −0.024 | 0.003 | <0.0001 |
| Caucasian | 0 | . | . |
| Gender | | | |
| Female | 0.111 | 0.001 | <0.0001 |
| Male | 0 | . | . |
| Diagnoses | | | |
| Diabetes | 0.233 | 0.001 | <0.0001 |
| Nondiabetes | 0 | . | . |
| Ethnicity | | | |
| Hispanic | 0.171 | 0.001 | <0.0001 |
| Non-Hispanic | 0 | . | . |
| Nursing home status | | | |
| Yes | 1.405 | 0.001 | <0.0001 |
| No | 0 | . | . |
| Log-BMI | −0.070 | 0.001 | <0.0001 |

## 6. Application

We applied the proposed marginal models with piecewise-constant baseline rates to the study of hospitalization days among Medicare dialysis patients. Between 2006 and 2008, there were 542,417 Medicare dialysis patients from 5650 dialysis facilities being hospitalized in the United States with facility sizes varying from 1 to 768 dialysis patients. The average hospitalization days per patient ranged from 1 to 788 with an average of 23 days during the 3 year period. In this study, hospital days are viewed as recurrent event data, with time of follow-up defined as time from 90 days after the initiation of ESRD therapy. We use the 90-day period to assure that most patients are eligible for Medicare insurance either as their primary or secondary insurer. Patients who died during the first 90 days of ESRD are excluded from the analysis. Patients are subject to left truncation at the start of the observation period, January 1, 2006. Subjects are followed until the earliest of death and right censoring, with the latter defined as the earliest of December 31, 2008, 3 days prior to transplant and loss to follow-up. Because a patient's hospitalization rate may be influenced by the facility at which (s)he receives dialysis, we fitted a facility-stratified model to adjust for facility effects.

Patient characteristics of interest include age, race, gender, diabetes, ethnicity, nursing home status and body mass index (BMI). All covariates except for BMI are coded as categorical variables through binary indicators. We include logarithm of BMI as a continuous covariate. According to the proposed methods, we summarize patient hospital days as intermittent counts and exposure times in six time-since-ESRD intervals with cut points 90 days (time 0), 6 months, 1 year, 2 years, 3 years, and 5 years. These cut points are selected based on previous knowledge that more frequent hospitalization could be observed at the early stages of ESRD therapy. Patient age is recorded at the beginning of each interval. Nursing home status is recorded as whether a patient was in nursing home in the previous calendar year. All the rest variables are measured at the beginning of the study, thus are time independent. Results of the covariate effects from the proposed method are summarized in Table 4 below.

All the included covariates significantly influence the recurrent rate of hospitalization days. When comparing within the same cluster, patients at the ages 45–59 have the lowest hospitalization rates among survivors with all the other patient mix held the same. Asian dialysis patients are less frequently hospitalized among survivors than the corresponding comparable groups. According to the model, on any given day, female patients are 1.12 times more likely to be in hospital than male patients. Similarly, diabetic patients are 1.26 times more likely to be in hospital than nondiabetic patients and Hispanic patients are 0.84 times less likely to be hospitalized than non-Hispanic patients. Conditional on being alive, dialysis patients who were in a nursing home in the previous calendar year are 4.01 times more likely to be hospitalized than patients who did not reside in a nursing home. For each unit increase in the logarithm of BMI, the log hazard of hospitalization decreases by 0.933.

Using the stratified model, we are also interested in the trend of facility-specific baseline rate functions. Because the

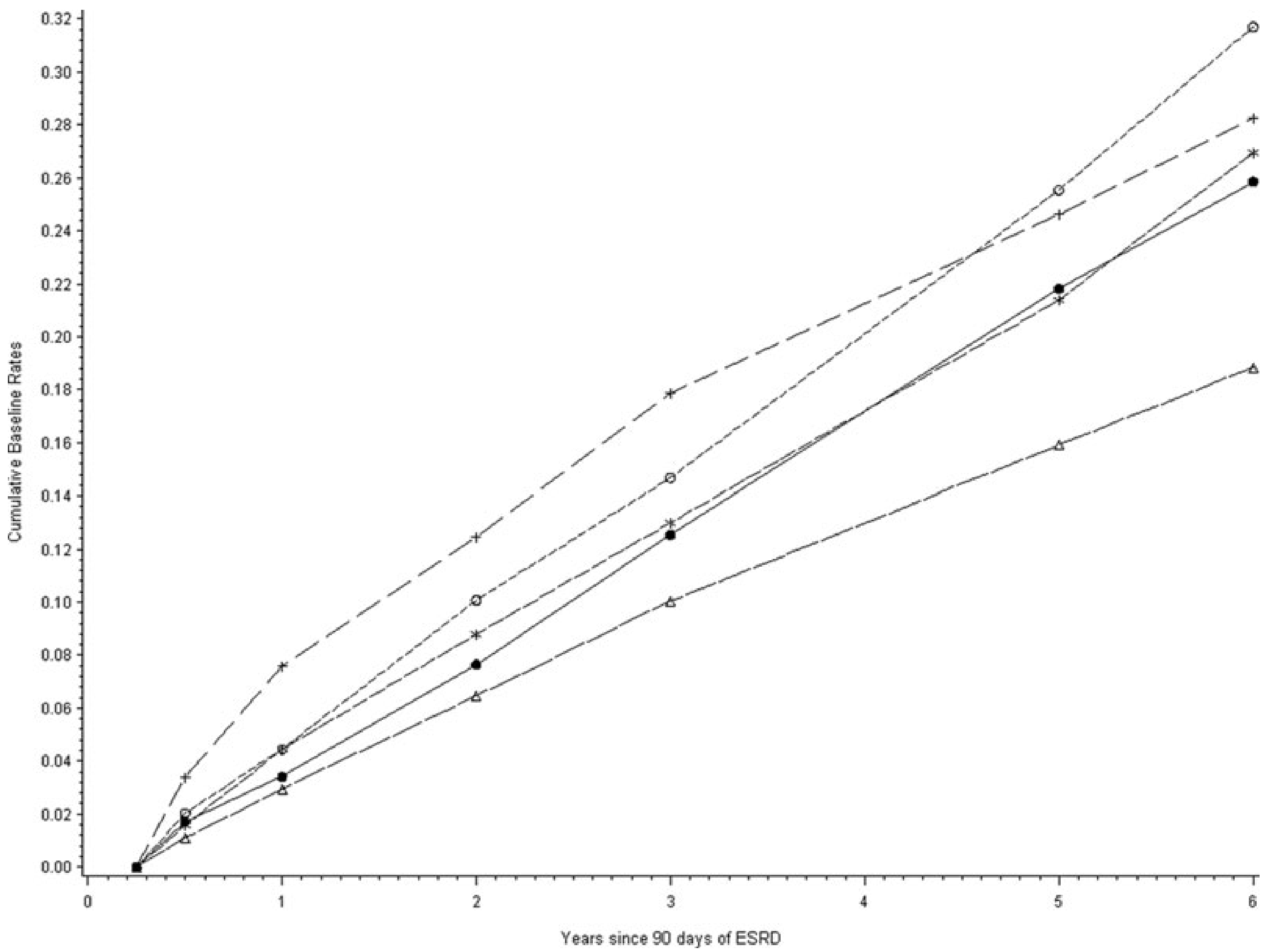


**Figure 1.** The cumulative baseline rates estimators for five facilities with more than 500 dialysis patients. The slopes within each interval represent the piecewise-constant baseline rates.

performance of the baseline rates estimators are improved with increasing center size, we choose the five largest facilities with more than 500 dialysis patients and estimate their baseline rates. The corresponding estimators are plotted in Figure 1. For all five facilities, the hospitalization baseline rates decrease with time on dialysis and this tendency is more obvious in the first year of dialysis treatment. After 1 year of ESRD therapy, the hospitalization rates are quite flat. This finding is coincident with our previous knowledge that dialysis patients are more likely to be hospitalized in early stage of ESRD therapy.

## 7. Discussion

In this article, we propose a proportional rates model with cluster-specific piecewise-constant baseline rate function for recurrent event data, which applies to the settings with and without a terminal event. With the parametric setting for the baseline rate function, we are able to estimate the regression parameter and cumulative baseline rates based on intermittent counts and exposure times within each prespecified interval, which is defined according to the pieces in the baseline rate function. The proposed method reduces data storage volume and speeds up the computation. The Cox format of the estimating equation enables the feasibility of stratification, which is difficult to implement under the joint estimating equation approach when the number of clusters is relatively large, as in the illustrating example in Section 6.

The proposed model is still a proportional rates model. General method for the checking of goodness of fit for the common Cox model can be applied directly. For example, to check for nonproportionality, one may consider including a nonlinear term for time and test the corresponding coefficients.

The proposed model in the presence of terminal event does not specify the dependence between the terminal event and the recurrent event times. As long as the true partial marginal model is a proportional rates model, the proposed method gives reasonable estimates as was shown in the simulation study. In the simulation studies, the terminal event time $D$ is independent of the frailty $W$ capturing the dependence for within-subject events. Therefore, the dependence between $D$ and the recurrent event process $N^*(t)$ is completely through the covariate $Z$. If we assume $D$ also depends on $W$, the partial marginal model would become

$$E\{dN_{ik}^*(t)\,|\,\boldsymbol{Z}_i, D_i \geq t\} = f_W(t;\boldsymbol{Z}_i)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i}d\mu_{0k}(t).$$

Hence the estimators for the covariates influencing the distribution of $D$ ($\widehat{\beta}_1$ in the simulation study) would be biased. Therefore the proposed method applies to the situation that the dependence between $D$ and $N^*(t)$ is completely through either $W$ or $\boldsymbol{Z}$, but not both. This conclusion is also verified based on some simulation studies.

The proposed method is applicable to both recurrent event and failure time data from large registry study or large observational study such as claims data in insurance or hospitalization data. When the number of distinct event times is large, we can fold the data by recording the counts and exposure time in prespecified intervals and analyze the folded data using the proposed method.

## 8. Supplementary Materials

Tables referenced in Sections 5 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

## Acknowledgements

The authors thank Tempie Hulbert-Shearon and Valarie Ashby of the University of Michigan Kidney Epidemiology and Cost Center (KECC) for their assistance with the data analysis. This work was supported in part by contract 500-2006-00042C from the CMS and a grant 2R01 DK070869 from the National Institutes of Health.

## References

Allison, P. D. (2010). *Survival Analysis using* SAS*: A Practical Guide*, 2nd edition. North Carolina: SAS Publishing.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting process: A large sample study. *Annals of Statistics* **10,** 1100–1120.

Cook, R. and Lawless, J. (1997). Marginal analysis of recurrent events and a terminal event. *Statistics in Medicine* **16,** 911–924.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Ghosh, D. and Lin, D. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12,** 663–688.

Jacobsen, M. (1989). Existence and unicity of mles in discrete exponential family distribution. *Scandinavian Journal of Statistics* **16,** 335–349.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. New York: Wiley.

Lawless, J. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

Lawless, J. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *The Canadian Journal of Statistics* **26,** 549–565.

Lin, D., Wei, L., 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,** 771–730.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

Schaubel, D. and Cai, J. (2005). Semiparametric methods for clustered recurrent event data. *Lifetime Data Analysis* **11,** 405–425.

Schaubel, D. and Zhang, M. (2010). Estimating treatment effects on the marginal recurrentevent mean in the presence of a terminating event. *Lifetime Data Analysis* **16,** 451–477.

Sun, J. and Kalbfleisch, J. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5,** 279–290.

Sun, J. and Wei, L. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society, Series B* **62,** 293–302.

Thall, P. and Lachin, J. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association* **83,** 339–347.

Wellner, J. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of statistics* **28,** 779–814.

Ye, Y., Kalbfleisch, J., and Schaubel, D. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63,** 78–87.

*Received August* 2010. *Revised July* 2011.
*Accepted July* 2011.

## Appendix

*Proof of Theorem 1*

Define

$$P_n(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\{\boldsymbol{\beta}^T\boldsymbol{Z}_{i\ell} - \log S_{k\ell}^{(0)}(\boldsymbol{\beta})\}d_{ik\ell},$$

and let $W_n(\boldsymbol{\beta}) = P_n(\boldsymbol{\beta}) - P_n(\boldsymbol{\beta}_0)$, which can be written as

$$W_n(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left\{(\boldsymbol{\beta}^T - \boldsymbol{\beta}_0^T)\boldsymbol{Z}_{i\ell} - \log\frac{S_{k\ell}^{(0)}(\boldsymbol{\beta})}{S_{k\ell}^{(0)}(\boldsymbol{\beta}_0)}\right\}d_{ik\ell}.$$

With condition (a) to (e) in Section 4, the strong law of large number and the fact that $d_{ik\ell}$ and $S_{k\ell}^{(0)}(\boldsymbol{\beta})$ have bounded variation, we can show that $W_n(\boldsymbol{\beta})$ converges almost surely to

$$\mathcal{W}(\boldsymbol{\beta}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L} \times\left\{\left(\boldsymbol{\beta}^T - \boldsymbol{\beta}_0^T\right)\boldsymbol{Z}_{i\ell} - \log\frac{s_{k\ell}^{(0)}(\boldsymbol{\beta})}{s_{k\ell}^{(0)}(\boldsymbol{\beta}_0)}\right\}\alpha_{k\ell}t_{ik\ell}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_{i\ell}},$$

for every $\boldsymbol{\beta}$. Obviously,

$$\begin{aligned}\frac{\partial^2 W_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} &= -n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\boldsymbol{V}_{k\ell}(\boldsymbol{\beta})d_{ik\ell},\\ &= -n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\{\boldsymbol{Z}_{i\ell} - \overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta})\}^{\otimes 2}\\ &\qquad\times d_{\bullet k\ell}/S_{k\ell}^{(0)}(\boldsymbol{\beta})t_{ik\ell}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_{i\ell}},\end{aligned}$$

is negative semidefinite. Therefore, $W_n(\boldsymbol{\beta})$ is concave. By Theorem 10.8 of Rockafellar (1970), the convergence of $W_n(\boldsymbol{\beta})$ to $\mathcal{W}(\boldsymbol{\beta})$ is uniform on any compact set of $\boldsymbol{\beta}$. Specifically, defining a compact set of $\boldsymbol{\beta}$, $\mathcal{B}_r = \boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq r$, we have

$$\sup\nolimits_{\boldsymbol{\beta}\in\mathcal{B}_r}\|W_n(\boldsymbol{\beta}) - \mathcal{W}(\boldsymbol{\beta})\| \to 0. \tag{A1}$$

In addition, $\partial \mathcal{W}(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta} = 0$ and $\partial^2 \mathcal{W}(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T = -A$, which is assumed to be negative semidefinite through condition ($g$). Hence, $\mathcal{W}(\boldsymbol{\beta})$ has a unique maximizer at $\boldsymbol{\beta}_0$. In particular, $\sup_{\beta \in \partial \mathcal{B}_r} \{\mathcal{W}(\boldsymbol{\beta})\} < \mathcal{W}(\boldsymbol{\beta}_0)$, where $\partial \mathcal{B}_r = \boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = r$ is the boundary of $\mathcal{B}_r$. This fact, together with expression (A1), implies that $W_n(\boldsymbol{\beta}) < W_n(\boldsymbol{\beta}_0)$ for all $\boldsymbol{\beta} \in \partial \mathcal{B}_r$ and large $n$. Therefore, there must exists a maximizer of $W_n(\boldsymbol{\beta})$, i.e. the solution to $\partial W_n(\boldsymbol{\beta})\ \partial \boldsymbol{\beta} = 0$, say $\widehat{\boldsymbol{\beta}}$, in the interior of $\mathcal{B}_r$, and the argument in Jacobsen (1989) can be used to show the uniqueness of this maximizer. Because $r$ can be arbitrarily small, letting $r \to 0$ yields that $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$ as $n \to \infty$.

The weak convergence of $\widehat{\boldsymbol{\beta}}$ can be shown through the first-order Taylor's series expansion about $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ on $\boldsymbol{U}(\widehat{\boldsymbol{\beta}})$ as

$$\boldsymbol{U}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{U}(\boldsymbol{\beta}_0) + \tfrac{\partial \boldsymbol{U}(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}^*$ is on the line segment joining $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. It follows that

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \widehat{\boldsymbol{A}}^{-1}(\boldsymbol{\beta}^*) n^{-1/2} \boldsymbol{U}(\boldsymbol{\beta}_0).$$

The almost sure convergence of $\widehat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}_0$ and the fact that $\xi_{ik\ell}(\boldsymbol{\beta}_0)$ is zero mean implies that $\widehat{\boldsymbol{A}}(\boldsymbol{\beta}^*) \xrightarrow{a.s.} \boldsymbol{A}$ as $n \to \infty$.

Next, we derive the distribution of $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0)$ beginning by the following decomposition

$$\begin{aligned} \boldsymbol{U}(\boldsymbol{\beta}_0) = & \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{\ell=1}^{\ell} \{\boldsymbol{Z}_{i\ell} - \bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta}_0)\} \xi_{ik\ell}(\boldsymbol{\beta}_0) \\ & - \sum_{k=1}^{K} \sum_{\ell=1}^{\ell} \{\overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta}_0) - \bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta}_0)\} \xi_{\bullet k\ell}(\boldsymbol{\beta}_0). \end{aligned} \tag{A2}$$

The first term on the right-hand side of equation (A2) is a sum of $n$ independent and identically distributed random vectors with zero mean and finite variance. The second term on the right-hand side of equation (A2) is $o_p(n^{1/2})$ because $\overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta}_0) - \bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta}_0) \xrightarrow{p} 0$ and $\|n^{-1/2}\xi_{\bullet k\ell}(\boldsymbol{\beta}_0)\| = O(1)$ by the boundness conditions ($d$) and ($e$).

Thus $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0)$ converges weakly to a $p$-variate normal vector with mean 0 and covariance matrix $\boldsymbol{\Xi}(\boldsymbol{\beta}_0)$ by the multivariate central limit theorem. From Slutsky's theorem and the consistency of $\widehat{\boldsymbol{A}}(\boldsymbol{\beta}^*)$ to $\boldsymbol{A}$, $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a $p$-variate normal vector with mean 0 and covariance matrix $\Sigma$.

*Proof of Theorem 2*

We now derive the asymptotic properties for $\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}})$. The asymptotic results of $\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}})$ then directly applies by combining the results of $\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}})$ over $t$.

We first consider the following decomposition

$$\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}}) - \alpha_{k\ell} = \phi_1 + \phi_2, \tag{A2}$$

where $\phi_1 = \widehat{\alpha}_{k\ell}(\boldsymbol{\beta}_0) - \alpha_{k\ell}$, and $\phi_2 = \widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}}) - \widehat{\alpha}_{k\ell}(\boldsymbol{\beta}_0)$. We can write $\phi_1 = n^{-1}\xi_{\bullet k\ell}(\boldsymbol{\beta}_0)/S^{(0)}_{k\ell}(\boldsymbol{\beta}_0)$. The strong law of large number and condition ($f$) implies that $\phi_1 \xrightarrow{a.s.} 0$. By Taylor's series expansion,

$$\phi_2 = -n^{-1}\frac{d_{\bullet k\ell}\overline{\boldsymbol{Z}}_{k\ell}(\boldsymbol{\beta}_0)}{S^{(0)}_{k\ell}(\boldsymbol{\beta}_0)}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(n^{-1/2}).$$

By the boundness conditions ($d$) and ($f$), and the almost sure convergence of $\widehat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}_0$, $\phi_2 \xrightarrow{a.s.} 0$. The almost sure convergence of $\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}})$ to $\alpha_{k\ell}$ then follows. This result, together with (5) implies that $\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}})$ converges almost surely to $\mu_{0k}(t)$ uniformly in $t$.

Next, we prove the weak convergence of $n^{1/2}\{\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}}) - \alpha_{k\ell}\}$. With the previously derived arguments and condition (f),

$$n^{1/2}\phi_1 = n^{-1/2}\sum_{i=1}^{n}\frac{\xi_{ik\ell}(\boldsymbol{\beta}_0)}{s^{(0)}_{k\ell}(\boldsymbol{\beta}_0)} + o_p(1), \tag{A3}$$

$$n^{1/2}\phi_2 = -n^{-1/2}\sum_{i=1}^{n}\alpha_{k\ell}\bar{\boldsymbol{z}}_{k\ell}(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{U}_i(\boldsymbol{\beta}_0) + o_p(1). \tag{A4}$$

It then follows that

$$n^{1/2}\{\widehat{\alpha}_{k\ell}(\widehat{\boldsymbol{\beta}}) - \alpha_{k\ell}\} = n^{-1/2}\sum_{i=1}^{n}\psi_{ik\ell}(\boldsymbol{\beta}_0) + o_p(1).$$

This result, together with (5) implies that

$$\begin{aligned} & n^{1/2}\{\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}}) - \mu_{0k}(t)\} \\ & = n^{-1/2}\sum_{i=1}^{n}\sum_{\ell=1}^{L}\psi_{ik\ell}(\boldsymbol{\beta}_0)(a_\ell \wedge t - a_{\ell-1} \wedge t) + o_p(1), \end{aligned}$$

which converges to a zero-mean Gaussian process with covariance function $\psi_k(s,\ t)$.