Statistica Sinica **23** (2013), 1761-1774
doi:http://dx.doi.org/10.5705/ss.2012.147s

SEQUENTIAL ANALYSIS OF THE COX MODEL UNDER RESPONSE DEPENDENT ALLOCATION

Xiaolong Luo[1], Gongjun Xu[2] and Zhiliang Ying[2]

[1]*Celgene Corporation and* [2]*Columbia University*

Abstract: Sellke and Siegmund (1983) developed the Brownian approximation to the Cox partial likelihood score as a process of calendar time, laying the foundation for group sequential analysis of survival studies. We extend their results to cover situations in which treatment allocations may depend on observed outcomes. The new development makes use of the entry time and calendar time along with the corresponding σ-filtrations to handle the natural information accumulation. Large sample properties are established under suitable regularity conditions.

Key words and phrases: Survival analysis, group sequential methods, outcome dependent allocation, proportional hazards regression, clinical trials, staggered entry, Brownian approximation, weak convergence.

1. Introduction

The Cox (1972) proportional hazards model along with the partial likelihood (Cox (1975)) has been extensively applied to survival data. The theoretical properties of the maximum partial likelihood estimator can be easily derived by expressing the partial likelihood score as a counting process-based martingale integral; see Andersen and Gill (1982), Fleming and Harrington (1991), and Kalbfleisch and Prentice (2002).

For sequential analysis, the partial likelihood score needs to be evaluated along the calendar time and its asymptotic behavior is crucial to deriving the corresponding group sequential methods. Due to the staggered entry of patients, the partial likelihood score as a process of calendar time is no longer a martingale integral. In a pioneering paper, Sellke and Siegmund (1983) showed that the score process can still be approximated by the Brownian motion process, thereby laying the foundation for group sequential analysis of survival studies. Slud (1984) also established the Brownian approximation to the log-rank process for survival outcome under staggered entry. A Gaussian random field approximation to the two-dimensional score process in the case of two-sample comparison was established by Gu and Lai (1991); see also Chapter 10 in Andersen et al. (1993). More general results about Gaussian random field approximation to the two-dimensional score process under the Cox proportional hazards regression can be

found in Bilias, Gu, and Ying (1997), where modern empirical process theory is applied to derive certain key results, bypassing the martingale formulation.

The results of Sellke and Siegmund (1983) can be readily applied in the context of group sequential analysis as described in Pocock (1977), O'Brien and Fleming (1979), and Lan and DeMets (1983). However, their results are not applicable under adaptive designs where treatment allocation may depend on preceding outcomes. This is because the outcome variables are dependent so that neither the counting process-martingale argument nor the empirical process theory may be used to derive the desirable Brownian motion approximation. For some initial ideas of adaptive design, see Thompson (1933) and Robbins (1952); for early works, see Zelen (1969), Wei and Durham (1978), and Wei (1978); for more recent developments, see Flournoy and Rosenberger (1995) and Hu and Rosenberger (2006).

The existing literature on response adaptive treatment allocation methods primarily deals with continuous or binary outcome variable. Recently Zhang and Rosenberger (2007) developed a parametric approach to survival outcomes. They assumed that survival times follow the exponential or, more generally, the Weibull family of distributions. They showed that their approach can result in approximately optimal treatment allocation assuming survival times are relatively shorter than the follow up period.

The main focus of this paper is to extend the results of Sellke and Siegmund (1983) to the situation in which treatment allocations may depend on preceding outcomes. A key step in the new development is the expression of the partial likelihood score process in terms of integrals over the calendar and entry times. As a result, the usual martingale structure is preserved and can be applied to establish large sample properties. Indeed, it is shown that the partial likelihood score process is approximated by a time-rescaled Brownian motion process and that the maximum partial likelihood estimator is asymptotically normal.

The remainder of this paper is organized as follows. Section 2 first explains why the current martingale approach fails under the outcome dependent allocations, and then introduces a new approach. The corresponding functional central limit theorems are presented in Section 3, where convergence properties for the corresponding maximum partial likelihood estimator are also established. Some discussion is in Section 4. Most technical developments are presented in the supplementary material.

2. Notation and Model Specification

We first introduce the setup and define some basic quantities. We consider a follow up study with calendar time period $[0, \tau]$, where $\tau < \infty$. Let n be the sample size of the study. Denote by $U_{n,i}$ the entry time for individual i, $i \geq 1$. For technical convenience, we assume that the $U_{n,i}$ have no ties. Thus, without

loss of generality, $U_{n,1} < U_{n,2} < \cdots < U_{n,i} < \cdots$. Define the associated counting process for entry times

$$R_n(t) = \sum_{i \geq 1} I_{(U_{n,i} \leq t)} \ , \tag{2.1}$$

so $R_n(t)$ is the total enrollment up to time t and $R_n(\tau) = n$. By large sample, we mean that n goes to infinity while τ remains fixed. In other words, the situation considered here is high rate of entry over a fixed time period. An example of such kind in survival studies is the Beta-Blocker Heart Attack Trial (BHAT (1982)), where 3837 persons entered during the 27-month follow up period. For notional convenience we omit the subscript n in $U_{n,i}$ when no confusion arises.

For subject i, let T_i denote the survival time (since entry) and C_i the censoring time. Throughout, $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, $a^+ = \max\{0, a\}$, and $a^- = \max\{0, -a\}$. Let $\tilde{T}_i = T_i \wedge C_i$ and $\Delta_i = I_{(T_i \leq C_i)}$, indicating failure (1) or censoring (0). Thus, if $\Delta_i = 1(0)$, then individual i experiences failure (censoring) at calendar time $U_i + \tilde{T}_i$. Furthermore, there is a p-dimensional covariate vector Z_i that may include ith individual's treatment assignment and certain relevant baseline characteristics.

We describe the Cox model specification with independent censoring under outcome dependent allocation as follow. For the ith subject, given Z_i, T_i is conditionally independent of C_i and $\{T_j, C_j, Z_j; j < i\}$ and has a proportional hazards model specification

$$\lambda_i(t) = \exp(\beta' Z_i)\lambda_0(t),$$

where β is an unknown p-dimensional regression parameter of interest and λ_0 is the baseline hazard function. Note that under adaptive allocation, given Z_j, T_j may not be independent of T_i if $i > j$ since Z_i, which includes the treatment allocation of the ith subject, may depend on the survival experiences of other subjects who enrolled before time U_i. For instance, in Figure 1, we can see that Z_i (and T_i) may depend on the survival information T_j under the outcome dependent allocation scheme. Compared with the independent enrollment scheme, as in Sellke and Siegmund (1983) where $\{T_i, C_i, Z_i\}$ are all assumed to be independent, outcome dependent allocation violates the independent assumption, raising the issue of validity for the existing sequential testing procedures. We demonstrate the theoretical challenges of the violation of independence in the next subsection, and propose our new approach in Subsection 2.2.

2.1. Partial likelihood score process over survival time

With the usual nonadaptive allocation, i.e., observations from individual units are mutually independent, the partial likelihood (Cox (1975)) takes the

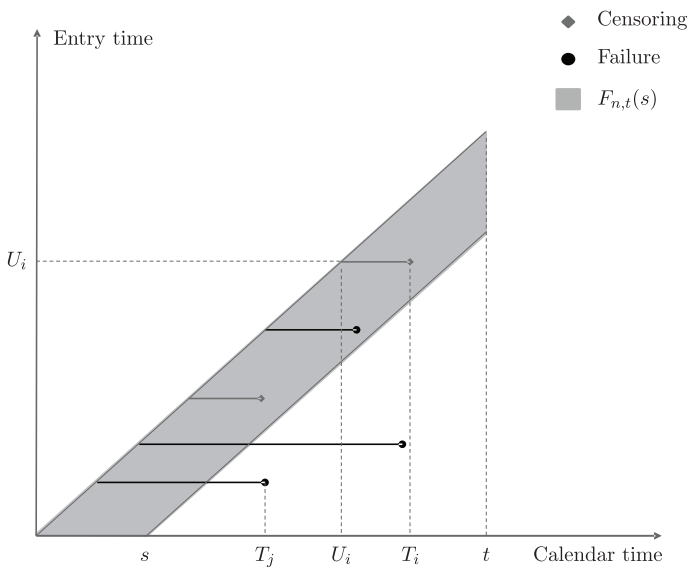

Figure 1. $\mathcal{F}_{n,t}(s)$: σ-filtrations defined along calendar time and survival time.

form

$$PL(t) = \prod_{\substack{i:\, \tilde{T}_i \le (t-U_i)^+, \\ \Delta_i = 1}} \left\{ \frac{\exp(\beta' Z_i)}{\sum_{j:\, \tilde{T}_j \le (t-U_j)^+, \ \tilde{T}_j \ge T_i} \exp(\beta' Z_j)} \right\}. \tag{2.2}$$

Taking logarithm and differentiating with respect to β results in the corresponding partial likelihood score process

$$U_n(t) = \sum_{i:\, U_i \le t} \int_0^t [Z_i - \bar{Z}_n(\beta; t, s)] N_i(t, ds), \tag{2.3}$$

where

$$\bar{Z}_n(\beta; t, s) = \frac{\sum_{i:\, U_i \le t-s} Z_i \exp(\beta' Z_i) I_{(\tilde{T}_i \ge s)}}{\sum_{i:\, U_i \le t-s} \exp(\beta' Z_i) I_{(\tilde{T}_i \ge s)}},$$

$$N_i(t, s) = \Delta_i I_{(\tilde{T}_i \le s \wedge (t-U_i)^+)} .$$

Let

$$M_i(t, s) = N_i(t, s) - \int_0^s I_{(\tilde{T}_i \wedge (t-U_i)^+ \ge w)} \exp(\beta' Z_i) \lambda_0(w) dw. \tag{2.4}$$

It is well known that the partial likelihood score does not change numerically when the N_i are replaced by the M_i, so

$$U_n(t) = \sum_{i:\, U_i \le t} \int_0^t [Z_i - \bar{Z}_n(\beta; t, s)] M_i(t, ds). \tag{2.5}$$

The integration in (2.5) is with respect to survival time s. Under the usual independent sampling scheme, the M_i are martingales in survival time s with a suitably defined σ-filtration as in (2.6) below (Andersen et al. (1993)). Furthermore, the integrands are predictable, and $U_n(t)$ is a martingale integral with respect to survival time s. As a result, the martingale central limit theorem (Rebolledo (1980)) can be applied to obtain the normal (Brownian) approximation.

Under outcome dependent allocation, we show that the martingale (along survival time s) argument is no longer valid. For $s \geq 0$, let $\mathcal{F}_{n,t}(s)$ be the σ-filtration generated by observations up to survival time s and calendar time t,

$$\mathcal{F}_{n,t}(s) = \sigma\Big\{I_{(U_i\leq t)}, \quad U_i I_{(U_i\leq t)}, \quad Z_i I_{(U_i\leq t)}, \quad I_{(\tilde{T}_i\leq s\wedge(t-U_i)^+)}, \\ N_i(t,s), \quad \tilde{T}_i I_{(\tilde{T}_i\leq s\wedge(t-U_i)^+)}, \quad i=1,\cdots,n\Big\}. \tag{2.6}$$

Figure 1 illustrates the information accumulated along survival time. The grey trapezoid area shows the filtration $\mathcal{F}_{n,t}(s)$. From Figure 1, we can see that for the ith subject enrolled at time U_i, although its survival time is less than s, its treatment allocation (Z_i) depends on the outcome information of T_j, which is outside of $\mathcal{F}_{n,t}(s)$. Therefore, $M_i(t,s)$ may not be a martingale with respect to filtration $\mathcal{F}_{n,t}(s)$ under outcome dependent allocation. However, if $\{T_i, C_i, Z_i\}$ are all independent as is the case in Sellke and Siegmund (1983) and Gu and Lai (1991), the $M_i(t,s)$ are still $\mathcal{F}_{n,t}(s)$ martingales in s for any fixed t.

2.2. Calendar time based score process

We introduce a new way to represent the partial likelihood score so that a useful martingale structure will arise. This representation expresses the score process in terms of integrals over entry time and calendar time. Use of entry time instead of survival time is natural in terms of the information accumulation from data and the adaptive treatment allocation process.

With a slight abuse of notation, let $\tilde{T}_u$, Z_u, and Δ_u refer to $\tilde{T}_i$, Z_i, and Δ_i when $u = U_i$, which is well defined since the U_i are distinct for different i. Define a random counting measure

$$p_n(ds\, du) = I_{(u+\tilde{T}_u=s,\Delta_u=1)} dR(u)$$

that defines a bivariate counting process along both calendar time s and entry time u. It equals 1 if there exists a subject i such that $U_i = u$ and $T_i = s - u$; otherwise it equals 0. Based on this two-dimensional counting process, the Cox score in (2.3) can be rewritten as an integral with respect to both calendar time and entry time:

$$U_n(t) = \int_0^t \int_0^s [Z_u - \bar{Z}_n(\beta; t, s-u)] p_n(ds\, du), \tag{2.7}$$

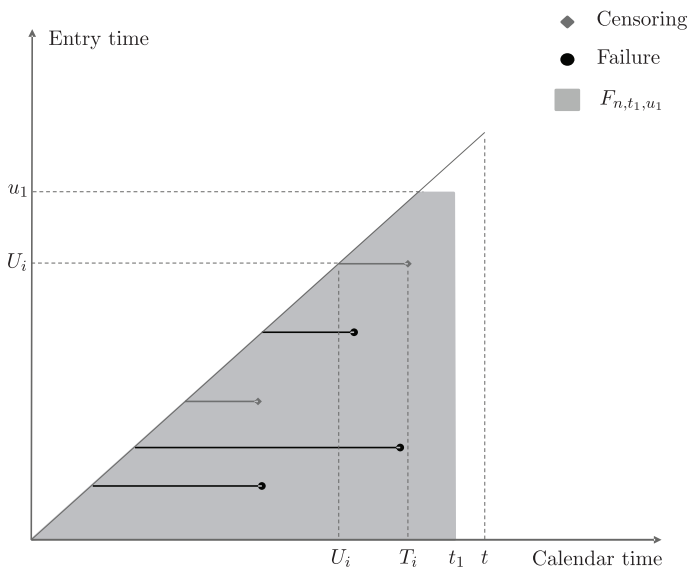

Figure 2. $\mathcal{F}_{n,t,u}$: σ-filtrations defined along calendar time and entry time.

Let $\mathcal{F}_{n,t}$ denote the corresponding σ-filtration containing all the information accumulation over calendar time period $[0,\ t]$,

$$\mathcal{F}_{n,t} = \sigma\Big\{I_{(U_i\le t)},\quad U_iI_{(U_i\le t)},\quad Z_iI_{(U_i\le t)},\\ I_{(\tilde{T}_i\le(t-U_i)^+)},\quad \Delta_iI_{(\tilde{T}_i\le(t-U_i)^+)},\quad \tilde{T}_iI_{(\tilde{T}_i\le(t-U_i)^+)};\quad i=1,\cdots,n\Big\}.$$

A sub-σ-algebra of $\mathcal{F}_{n,t}$ that is of interest is

$$\mathcal{F}_{n,t,\vartheta} = \sigma\Big\{I_{(U_i\le\vartheta)},\quad U_iI_{(U_i\le\vartheta)},\quad Z_iI_{(U_i\le\vartheta)},\quad I_{(\tilde{T}_i\le(t-U_i)^+,U_i\le\vartheta)},\\ \Delta_iI_{(\tilde{T}_i\le(t-U_i)^+,U_i\le\vartheta)},\quad \tilde{T}_iI_{(\tilde{T}_i\le(t-U_i)^+,U_i\le\vartheta)},\quad i=1,\cdots,n\Big\}.$$

Intuitively, $\mathcal{F}_{n,t,\vartheta}$ represents information up to calendar time t for individuals who enrolled before time ϑ, where $0<\vartheta\le t$. See Figure 2 for an illustration. The grey trapezoid area shows the filtration $\mathcal{F}_{n,t_1,u_1}$ that contains all the information up to calendar time t_1 and enrollment time u_1. Compared with Figure 1, we can see that the treatment allocation information of the ith subject (enrolled at time U_i) is now included in the new filtration.

Without loss of generality, we assume that $R(t)$ and Z_t are predictable with respect to $\{\mathcal{F}_{n,t}, t\ge 0\}$, which is standard in survival analysis. Note that for the ith subject, by the Dood-Meyer decomposition and the Cox model assumption, the compensator for the counting measure $p(ds, u = U_i) = I_{(U_i+\tilde{T}_i=s,\Delta_u=1)}$ is

$I_{(\tilde{T}_i \geq s-U_i)} \exp\{\beta' Z_i\}\lambda_0(s-U_i)ds$ when $s > U_i$. This follows from the fact that

$$P(U_i + T_i \in (s, s+ds]|\mathcal{F}_{n,s}) = I_{(\tilde{T}_i > s-U_i)} \exp\{\beta' Z_i\}\lambda_0(s-U_i)ds.$$

More generally, let

$$q_n(ds\, du) = I_{(\tilde{T}_u \geq s-u)} \exp\{\beta' Z_u\}\lambda_0(s-u)dR(u)ds.$$

Note that $I_{(u<s)}q_n(ds\, du)$ is the compensator of $I_{(u<s)}p_n(ds\, du)$. Thus we have the following lemma.

Lemma 1. *For $t \in (0, \tau]$,*

$$M_n(t) \triangleq \int_0^t \int_0^t I_{(u<s)}[p_n(ds\, du) - q_n(ds\, du)] \tag{2.8}$$

is a $\{\mathcal{F}_{n,t}, t \geq 0\}$ martingale. Moreover, for fixed t,

$$M_n(t, \vartheta) \triangleq \int_0^t \int_0^{\vartheta} I_{(u<s)}[p_n(ds\, du) - q_n(ds\, du)], \tag{2.9}$$

as a process in ϑ, is a $\{\mathcal{F}_{n,t,\vartheta}, 0 \leq \vartheta \leq t\}$ martingale.

Let $M_n(ds\, du) = I_{(u<s)}[p_n(ds\, du) - q_n(ds\, du)]$ be the corresponding martingale measure. The Cox score process in (2.7) can then be written as

$$\begin{aligned} U_n(\beta; t) &= \int_0^t \int_0^t [Z_u - \bar{Z}_n(\beta; t, s-u)] I_{(u<s)} p_n(ds\, du) \\ &= \int_0^t \int_0^t [Z_u - \bar{Z}_n(\beta; t, s-u)] M_n(ds\, du). \end{aligned}$$

More generally, we can define a two-parameter score process with respect to calendar time t and entry time ϑ as

$$U_n(\beta; t, \vartheta) = \int_0^t \int_0^{\vartheta} [Z_u - \bar{Z}_n(\beta; t, s-u)] M_n(ds\, du). \tag{2.10}$$

Note that $U_n(\beta; t, t) = U_n(\beta; t)$.

The expression here for $U_n(\beta; t)$ is an integral along the calendar time instead of the survival time as in standard counting process approach to survival analysis. Through this framework, responses and covariates history is expressed by the filtration $\mathcal{F}_{n,t}$. As a result, it is not difficult to show that $M_{n,t}$ is a martingale with respect to σ-filtration $\mathcal{F}_{n,t}$ (Lemma 1). This forms a crucial step for us to use the martingale central limit theorem to obtain the convergence for $U_n(\beta; t)$; see Section 3 for more details.

3. Large Sample Theory

In this section, we establish large sample properties that are important for the usual statistical inferences, especially for sequential analysis. We first deal with the score process, and then with the estimator.

3.1. Weak convergence of score process

We show here the weak convergence of U_n to a Gaussian random process. This extends results of Sellke and Siegmund (1983), Gu and Lai (1991), and Bilias, Gu, and Ying (1997) to cover the case with outcome dependent allocation schemes.

We adopt the setting of Bilias, Gu, and Ying (1997) and restrict t to $[0,\tau]$ with τ satisfying

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} P(\tilde{T}_i \geq \tau) > 0 \tag{3.1}$$

and λ_0 being bounded on $[0,\tau]$. Here we are adopting asymptotics in terms of a high rate of entry over a fixed time interval (large n), as opposed to a fixed rate of entry over a long time interval (large τ); see Siegmund (1985; p. 126). This allows us to develop a Gaussian random field approximation as in Gu and Lai (1991), which also assumes large n. For asymptotics under large τ, certain rescaling is needed and the corresponding Gaussian approximations can also be developed under certain stability assumptions (Siegmund (1985)).

For a p-dimensional covariate vector Z with regression parameter vector β, let $Z^{\otimes 0}=1$, $Z^{\otimes 1}=Z$, and $Z^{\otimes 2}=ZZ'$. For $k=0,1$ and 2, $\vartheta>0$, and $w>0$, let

$$\begin{aligned} S_{n,k}(\beta;\vartheta,w) &= \sum_{i:\, U_i\leq\vartheta} Z_i^{\otimes k}\exp(\beta' Z_i) I_{(\tilde{T}_i\geq w)} \\ &= \int_0^{\vartheta} I_{(\tilde{T}_u\geq w)} Z_u^{\otimes k}\exp(\beta' Z_u) dR(u). \end{aligned} \tag{3.2}$$

As in Section 2.2, take

$$\begin{aligned} U_n(\beta;t) &= \int_0^t \int_0^t [Z_u - \bar{Z}_n(\beta;t,s-u)] M_n(ds\, du), \\ U_n(\beta;t,\vartheta) &= \int_0^t \int_0^{\vartheta} [Z_u - \bar{Z}_n(\beta;t,s-u)] M_n(ds\, du), \end{aligned}$$

where

$$\bar{Z}_n(\beta;t,w) = \frac{S_{n,1}(\beta;t-w,w)}{S_{n,0}(\beta;t-w,w)}.$$

Let β_0 be the true regression parameter. We require conditions.

C1 The Z_i are uniformly bounded in the sense that there exists a non-random constant K_τ such that $\sup_{i:\, U_i \le \tau} |Z_i| \le K_\tau$, where $|\cdot|$ denotes the L_1 norm for a p-dimensional vector.

C2 For $k = 0$, 1, and 2, there exist non-random constants $\bar{E}_k(\vartheta, w)$ such that, as $n \to \infty$,

$$\frac{1}{n} S_{n,k}(\beta_0; \vartheta, w) - \bar{E}_k(\vartheta, w) \xrightarrow{P} 0$$

uniformly for all positive ϑ, w satisfying $\vartheta + w \le \tau$.

Remark 1. Conditions C1 and C2 are analogous to Conditions 1−3 in Bilias, Gu, and Ying (1997). In particular, C1 can be extended to a moment condition on Z for the components related to the baseline covariates. Condition C2 is required so that the sample moments for the Z_i are stable.

Theorem 1. *If* C1 *and* C2 *are satisfied, then*

(i) $n^{-1/2} U_n(\beta_0; t)$ *converges weakly to a vector-valued zero-mean Gaussian process* ξ *on* $[0, \tau]$ *with covariance function*

$$E[\xi(t_1)\xi'(t_2)] = \int_0^{t_1 \wedge t_2} \left[\bar{E}_2(t_1 \wedge t_2 - w, w) - \frac{\bar{E}_1^{\otimes 2}(t_1 \wedge t_2 - w, w)}{\bar{E}_0(t_1 \wedge t_2 - w, w)} \right] \lambda_0(w) dw;$$

(ii) $n^{-1/2} U_n(\beta_0; t, \vartheta)$ *converges weakly to a vector-valued zero-mean Gaussian random field* $\tilde{\xi}(t, \vartheta)$ *on* $\{(t, \vartheta) : 0 \le \vartheta \le t \le \tau\}$ *with covariance function*

$$\begin{aligned} &E[\tilde{\xi}(t_1, u_1)\tilde{\xi}'(t_2, u_2)] \\ &\quad = \int_0^{t_1 \wedge t_2} \left[\tilde{E}_2(u_1, u_2, t_1, t_2, w) - 2 \frac{\bar{E}_1(t_1 \wedge t_2 - w, w)}{\bar{E}_0(t_1 \wedge t_2 - w, w)} \tilde{E}_1'(u_1, u_2, t_1, t_2, w) \right. \\ &\qquad \left. + \frac{\bar{E}_1^{\otimes 2}(t_1 \wedge t_2 - w, w)}{(\bar{E}_0(t_1 \wedge t_2 - w, w))^2} \tilde{E}_0(u_1, u_2, t_1, t_2, w) \right] \lambda_0(w) dw, \end{aligned}$$

where $\tilde{E}_k(u_1, u_2, t_1, t_2, w) = \bar{E}_k(u_1 \wedge u_2 \wedge (t_1 \wedge t_2 - w), w)$, $k = 0, 1,$ *and* 2.

Remark 2. Theorem 1 extends existing results by allowing allocation schemes to be dependent on previous information. In addition, it implies that $\tilde{\xi}$ has independent increments in calender time t. Thus the diagonal process $\tilde{\xi}(t, t) = \xi(t)$ is a time-rescaled Brownian motion when $dim(Z) = 1$, and a vector-valued Gaussian process with independent increments when $dim(Z) > 1$.

Remark 3. To apply Theorem 1, we need to estimate the covariance function $E[\tilde{\xi}(t_1, u_1)\tilde{\xi}'(t_2, u_2)]$. A natural approach is to replace the unknown quantities $\bar{E}_k$ and $\Lambda(\cdot)$ with $S_{n,k}/n$ and the Nelson-Aalen estimator, respectively. Consistency of the corresponding covariance estimator can be derived under C1 and C2.

The proof of the next lemma, which plays a key role in the proof of Theorem 1, is given in the supplementary material.

Lemma 2. *Under the assumptions of Theorem 1,*

$$\sup_{\vartheta,t\in[0,\tau]} \frac{1}{\sqrt{n}} \int_0^t \int_0^\vartheta \left[\bar{Z}_n(\beta_0; t, s-u) - \frac{\bar{E}_1(t-(s-u), s-u)}{\bar{E}_0(t-(s-u), s-u)} \right] M_n(ds\, du) \xrightarrow{P} 0.$$

Remark 4. Lemma 2 shows that $\bar{Z}_n$ may be replaced by its (non-random) limit. The replacement makes it easy to use the martingale structure along the calendar time and the entry time without appealing to the empirical process theory that may not apply.

Proof of Theorem 1. When $\vartheta = t$, $U_n(\beta_0; t, \vartheta) = U_n(\beta_0; t)$, so we need only prove the weak convergence of $n^{-1/2} U_n(\beta_0; t, \vartheta)$. By Lemma 2, it suffices to show the weak convergence of

$$n^{-1/2} \tilde{U}_n(\beta_0; t, \vartheta) = n^{-1/2} \int_0^t \int_0^\vartheta \left[Z_u - \frac{\bar{E}_1(t-(s-u), s-u)}{\bar{E}_0(t-(s-u), s-u)} \right] M_n(ds\, du).$$

We first show that for any positive integer k and partition $0 \le u_1 < \cdots < u_k \le \tau$, $\{n^{-1/2}\tilde{U}_n(\beta_0; t, u_1), \ldots, n^{-1/2}\tilde{U}_n(\beta_0; t, u_k),\ 0 \le t \le \tau\}$ converges weakly to a multivariate Gaussian process $\{\tilde{\xi}(t, u_1), \ldots, \tilde{\xi}(t, u_k), 0 \le t \le \tau\}$. By Lemma 1, we have that the $\{\tilde{U}_n(\beta_0; t, u_j), \mathcal{F}_{n,t}, 0 \le t \le \tau\}$ are martingales along calendar time t with predictable variation processes

$$\begin{aligned}
&\langle n^{-1/2}\tilde{U}_n(\beta_0; \ldots, u_i), n^{-1/2}\tilde{U}_n(\beta_0; \ldots, u_j)\rangle(t) \\
&\quad = \frac{1}{n} \int_0^t \int_0^{u_i \wedge u_j \wedge s} \left[Z_u - \frac{\bar{E}_1(t-(s-u), s-u)}{\bar{E}_0(t-(s-u), s-u)} \right]^{\otimes 2} q_n(ds\, du) \\
&\quad \xrightarrow{P} \int_0^t \Bigg[\bar{E}_2(u_i \wedge u_j \wedge (t-w), w) - 2\frac{\bar{E}_1(t-w, w)}{\bar{E}_0(t-w, w)} \bar{E}_1'(u_i \wedge u_j \wedge (t-w), w) \\
&\qquad + \frac{\bar{E}_1^{\otimes 2}(t-w, w)}{(\bar{E}_0(t-w, w))^2} \bar{E}_0(u_i \wedge u_j \wedge (t-w), w) \Bigg] \lambda_0(w) dw,
\end{aligned}$$

where the convergence in probability is uniform in t and follows from C2. By the martingale central limit theorem (Rebolledo (1980)), any linear combination of $\{n^{-1/2}\tilde{U}_n(\beta_0; t, u_1), \ldots n^{-1/2}\tilde{U}_n(\beta_0; t, u_k), 0 \le t \le \tau\}$ converges weakly to the corresponding linear transformation of $\{\tilde{\xi}(t, u_1), \ldots, \tilde{\xi}(t, u_k), 0 \le t \le \tau\}$. Therefore, we obtain the weak convergence of $\{n^{-1/2}\ \tilde{U}_n(\beta_0; t, u_1), \ldots, n^{-1/2}\tilde{U}_n\ (\beta_0; t, u_k),$ $0 \le t \le \tau\}$ via the Cramér-Wold device. In particular, $n^{-1/2}\tilde{U}_n(\beta_0; t, \vartheta)$ converges in finite dimensional distributions to a Gaussian random field.

In the supplementary material, it is shown (Proposition 1) that for any $\epsilon > 0$, there exist a constant $k_0 < \infty$ and partition $0 = u_{n,0} \le u_{n,1} \le \cdots \le u_{n,k_0} = \tau$ such that, for all large n,

$$P\Big(\max_{0\le j<k_0} \sup_{\substack{\vartheta\in[u_{n,j},u_{n,j+1}];\\ 0\le t\le\tau}} \frac{1}{\sqrt{n}} |\tilde{U}_n(\beta_0;t,\vartheta) - \tilde{U}_n(\beta_0;t,u_{n,j})| \ge \epsilon \Big) \le \epsilon.$$

Thus, $n^{-1/2}\tilde{U}_n(\beta_0;t,\vartheta)$ is tight. Combined with the finite dimensional distributional convergence result, we obtain the desired conclusion.

3.2. Asymptotic normality of maximum partial likelihood estimator

We can use $U_n(\beta_0;t,\vartheta)$ to obtain an asymptotically unbiased estimator of β for each fixed (t,ϑ). Specifically, let $\hat{\beta}(t,\vartheta)$ be the solution to $U_n(\beta;t,\vartheta) = 0$. At $\vartheta = t$, $\hat{\beta}(t,t)$ is simply the maximum partial likelihood estimator with observable data at calendar time t. We show in this subsection that $\hat{\beta}(t,\vartheta)$ is asymptotically normal.

We first state a condition that ensures that the information matrix is nonsingular when normalized by the sample size n.

C3 There exists $\tau_0 \in (0,\tau]$ such that for all (ϑ,τ) satisfying $\tau_0 \le \vartheta \le t \le \tau$, $\lambda_{\min}(A(t,\vartheta)) \ge v_0 > 0$, *a.s.*, where

$$A(t,\vartheta) = \int_0^t \left[\bar{E}_2(t-w,w) - \frac{\bar{E}_1^{\otimes 2}(t-w,w)}{\bar{E}_0(t-w,w)} \right] \frac{\bar{E}_0(\vartheta\wedge(t-w),w)}{\bar{E}_0(t-w,w)} \lambda_0(w)dw,$$

$\bar{E}_k$ defined as in C2 and $\lambda_{\min}(A)$ the minimum eigenvalue of a symmetric matrix A.

Theorem 2. *Suppose that* C1*,* C2*, and* C3 *are satisfied. Then* $\{\sqrt{n}(\hat{\beta}(t,\vartheta) - \beta_0), \tau_0 \le \vartheta \le t \le \tau\}$ *converges weakly to a vector-valued zero-mean Gaussian process* η *with covariance function*

$$E[\eta(t_1,u_1)\eta'(t_2,u_2)] = (A(t_1,u_1))^{-1} E[\tilde{\xi}(t_1,u_1)\tilde{\xi}'(t_2,u_2)] (A(t_2,u_2))^{-1},$$

where $\tilde{\xi}$ *is the Gaussian process defined as in Theorem* 1*.*

Proof of Theorem 2. By Lemma 3 in the supplementary material, we have that, as $n \to \infty$,

$$\sup_{0\le\vartheta\le t\le\tau} \left| \frac{1}{n} U_n(\beta_0;t,\vartheta) \right| \xrightarrow{P} 0. \tag{3.3}$$

Condition C2 implies that

$$\sup_{0\le\vartheta\le t\le\tau} \left| \frac{1}{n}\frac{\partial}{\partial\beta} U_n(\beta_0;t,\vartheta) + A(t,\vartheta) \right| \xrightarrow{P} 0. \tag{3.4}$$

Since $\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta_0;t,\vartheta)$ has a uniformly bounded derivative with respect to β, C3 and (3.4) imply that there exists a neighborhood of β_0, $\mathcal{N}(\beta_0)$, such that

$$\liminf_{n\to\infty}\inf_{\tau_0\le\vartheta\le t\le\tau}\inf_{\beta\in\mathcal{N}(\beta_0)}\lambda_{\min}\Big(-\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta;t,\vartheta)\Big)\ge\frac{v_0}{2}>0. \tag{3.5}$$

Therefore, by (3.3), (3.5), and Lemma 5 in the supplementary material, together with the positive definiteness of $-\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta;t,\vartheta)$, we have

$$\sup_{\tau_0\le\vartheta\le t\le\tau}|\hat\beta(t,\vartheta)-\beta_0|\xrightarrow{P}0.$$

By the Taylor series expansion, we have that

$$\begin{aligned}0&=\frac{1}{\sqrt{n}}U(\hat\beta(t,\vartheta);t,\vartheta)\\&=\frac{1}{\sqrt{n}}U(\beta_0;t,\vartheta)+\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta_0;t,\vartheta)\sqrt{n}(\hat\beta(t,\vartheta)-\beta_0)+o_p(1),\end{aligned}$$

uniformly in $\tau_0\le\vartheta\le t\le\tau$. Therefore

$$\sqrt{n}(\hat\beta(t,\vartheta)-\beta_0)=-\left(\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta_0;t,\vartheta)\right)^{-1}\frac{1}{\sqrt{n}}U(\beta_0;t,\vartheta)+o_p(1).$$

The weak convergence of $\sqrt{n}(\hat\beta(t,\vartheta)-\beta_0)$ follows from this expansion and Theorem 1.

4. Discussion

One of the limitations of the asymptotic theory developed here is the assumption of high accrual rate in a fixed follow up period. Such an assumption entails that a significant portion of survival experiences from previously entered subjects may not be fully available for optimal treatment allocation due to delayed survival outcomes. Consequently, the asymptotically optimal treatment allocation ratio as discussed in Zhang and Rosenberger (2007) may not be attainable. On the other hand, the flexibility of using all observed survival outcomes could alleviate this deficiency of delayed response.

Another way to formulate large sample setting is to assume large time, rather than high accrual rate, so τ (follow up period) goes to infinity. Under this formulation for large (calendar time) t, the proportion of observed outcomes from previously entered subject will tend to 1 as t goes to infinity, making the asymptotically optimal treatment allocation feasible. When there is no other explanatory variable besides a dichotomous treatment allocation, it is not difficult to extend the present approach by rescaling of time through the "compensator".

In general, it may require additional assumptions on the explanatory variables in order to establish the vector-valued Gaussian martingale approximation to the multivariate score process.

To alleviate the effect of delayed survival outcomes, certain surrogate variables (markers) for the survival time may be used for the purpose of treatment allocation. For example, in the BATTLE trial (Zhou et al. (2008); Kim et al. (2011)), if patients' survival times were the endpoint, then one could use progression-free survival as a surrogate variable. It is of interest to develop a similar theoretic framework under which the Brownian approximation may be used.

The approach developed here may be extended to other follow-up studies with more general outcome variables. For studies with longitudinal outcomes, dynamic regression models have been proposed and studied (Martinussen and Scheike (2000)). Adaptive and outcome dependent designs for such studies may result in staggered entry and dependent observation units. We believe the general approach developed in this paper can be extended to deal with such designs.

Acknowledgement

The authors are grateful to an associate editor and two referees for their comments and suggestions which led to many clarifications and a better and more focused presentation. This research was supported in part by the NIH grant R37GM047845.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Andersen, P. K. and Gill, R. D. (1982). Coxs regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

BHAT (1982). A randomized trial of propranolol in patients with acute myocardial infarction. *J. Amer. Med. Assoc.* **147**, 1707-1714.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25**, 662-682.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

Fleming, T. R. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Flournoy, N. and Rosenberger, W. F. (1995). *Adaptive Designs*. IMS, Hayward, CA.

Gu, M. G. and Lai, T. L. (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Ann. Statist.* **19**, 1403-1433.

Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-adaptive Randomization in Clinical Trials*. Wiley, New York.

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

Kim, E. S., et al. (2011). The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discovery* **1**, 44-53.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659-663.

Martinussen, T. and Scheike, T. H. (2000). A nonparametric dynamic additive regression model for longitudinal data. *Ann. Statist.* **28**, 1000-1025.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549-556.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**, 191-199.

Rebolledo, R. (1980). Central limit theorem for local martingales. *Z. Wahrsch. Verw. Gebiete* **51**, 269-286.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58**, 527-535.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70**, 315-26.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

Slud, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. *Ann. Statist.* **12**, 551-571.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of the two samples. *Biometrika* **25**, 285-294.

Wei, L. J. (1978). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6**, 92-100.

Wei, L. J. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73**, 840-843.

Zelen, M. (1969). Play the winner and the controlled clinical trial. *J. Amer. Statist. Assoc.* **64**, 131–146.

Zhang, L. and Rosenberger, W. F. (2007). Response-adaptive randomization for survival trials: the parametric approach. *Appl. Statist.* **56**, 153-165.

Zhou, X., Liu, S., Kim, E. S., Herbst, R. S. and Lee, J. J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer - a step toward personalized medicine. *Clinical Trials* **5**, 181-193.

Celgene Corporation, 86 Morris Avenue Summit, NJ 07901, USA.

E-mail: xluo@celgene.com

Department of Statistics, Columbia University, MC 4690, 1255 Amsterdam Avenue New York, NY 10027, USA.

E-mail: gongjun@stat.columbia.edu

Department of Statistics, Columbia University, MC 4690, 1255 Amsterdam Avenue New York, NY 10027, USA.

E-mail: zying@stat.columbia.edu

(Received May 2012; accepted May 2013)