Temporal Process Regression

Jun Yan,[1,*] **Jason P. Fine**[1,2] **and Michael. R. Kosorok**[1,2]

[1]Department of Statistics, University of Wisconsin–Madison

1210 West Dayton St., Madison, WI USA

[2]Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison

600 Highland Ave., Madison, WI USA

Summary. We consider regression for response and covariates which are temporal processes observed over intervals. A functional generalized linear model is proposed which includes extensions of standard models in multistate survival analysis. Simple nonparametric estimators of time-indexed parameters are presented and shown to be uniformly consistent and to converge weakly to Gaussian processes. The procedure does not require smoothing or a Markov assumption, unlike approaches based on transition intensities. The estimators are the basis for new tests of the covariate effects and for the estimation of models in which greater structure is imposed on the parameters. The methodology enables goodness-of-fit testing and permits predictions involving estimated components from both the functional model and the submodels. Its practical utility is illustrated in recurrent event simulations and a data analysis of the prevalence of Chronic Graft Versus Host Disease (CGVHD).

* *email:* jyan@stat.wisc.edu

Key words: Empirical process; Functional estimating equation; Partially observed; Prevalence; Recurrent event; Uniform convergence; Varying coefficient

1. Introduction

Hazard regression with time-dependent parameters is a useful exploratory analysis for covariates with right censored data. With complex multistate data and other observational schemes, the intensity may be less attractive for reasons of interpretation and estimation. Consider bone marrow transplantation studies in which the effect of prophylaxis on the prevalence of graft versus host disease may be of scientific interest (Pepe et al., 1991). Regression analyses based on transition intensities yield indirect information on the quantity. The issue is that the probability is derived from multiple intensities, each with a model having time-dependent coefficients.

We propose an alternative, the functional generalized linear model. The mean of a response $Y(t)$ at time t is specified conditionally on a $p \times 1$ vector of time-dependent covariates $X(t)$ and a time-dependent stratification factor $S(t)$. That is,

$$E\{Y(t)|X(t), S(t)\} = g^{-1}\{\beta^T(t)X(t)\} \tag{1}$$

where the link function g is monotone, differentiable, and invertible, and $\beta(t) = \{\beta_1(t), \ldots, \beta_p(t)\}^T$ is a $p \times 1$ vector of time-dependent coefficients for the strata under consideration. The parameter $\beta(t)$ has a clear meaning in the model at t and because the link is time-independent, $\beta(s)$ and $\beta(\tilde{s})$ are comparable for $s \neq \tilde{s}$. In the bone marrow example, taking $g^{-1}(u) = \exp(u)/\{1 + \exp(u)\}$ gives a time-indexed logistic model with $\beta(t)$ denoting the change in the log odds ratios for graft versus host disease per unit increases in the covariates at time t.

In practice, the data processes may be missing at some times. Let $R(t)$ equal 1 if $\{Y(t), X(t), S(t)\}$ is fully observed at t, and 0 if not. Assume that for fixed t, $Y(t)$ and $R(t)$ are independent conditionally on $\{X(t), S(t)\}$. The set-up is similar to that in Nadeau and Lawless (1998). It includes many scenarios where model (1) with parametric $\beta(t)$ is the standard analysis. One is Pepe, Heagerty and Whitaker (1999) model for the prevalence function (Pepe and Couper, 1997). Another is Andersen and Gill (1982) model for recurrent event data, generalized to transformation models by Lin, Wei and Ying

(2001). A third is Lin (2000) proportional means model for medical costs data. The focus of this paper is different from these earlier works, in that the coefficients are completely unspecified.

Nonparametric inference for time-dependent coefficients has been well studied in proportional hazards regression (Zucker and Karr, 1990; Murphy and Sen, 1991; Fahrmeir and Klinger, 1998) and the additive hazards model (Aalen, 1980; Huffer and McKeague, 1991; McKeague and Sasieni, 1994). Typically, estimation involves smoothing over time and is complicated. In our formulation, the probability of failure by t is modeled conditionally on $\{X(t), S(t)\}$, instead of the hazard at t being modeled conditionally on $\{X(s), S(s), s \leq t\}$. The Markov structure of the intensity is not required for model (1) to hold. That is, $E\{Y(t)|X(t), S(t)\}$ may not equal $E\{Y(t)|X(s), S(s), s \leq t\}$. This specification is convenient for estimation.

We exploit that the model (1) only posits the conditional mean of $Y(t)$, not its covariance. Moment methods (Liang and Zeger, 1986) which do not restrict the processes' temporal dependence are adapted to separately estimate $\beta(t)$ at each time point, without smoothing. In Section 2, an estimating equation is presented which leads to simple nonparametric estimators. Their pointwise properties follow from existing results. A challenge is establishing that the theory holds uniformly in t. Since we model means and not intensities, martingale theory (Andersen et al., 1995) is not applicable and empirical processes (van der Vaart and Wellner, 1996) are needed. The arguments are provided in the appendix. The uniform convergence is essential for the developments in later sections.

Conceptually, the modeling strategy is functional data analysis (Ramsay and Dalzell, 1991). It is related to varying-coefficient models (Hastie and Tibshirani, 1993) for longitudinal data at finite irregularly placed times. In Lin and Ying (2001) inferences for the integrated coefficients from a linear regression are developed without smoothing. However, as with hazards models, most approaches to discretely observed data with $g(u) = u$ require local estimation (Hoover et al., 1998; Wu et al., 1998; Martinussen and Scheike, 1999; Fan and Zhang, 2000). What distinguishes our data is that it is

observed continuously on intervals, which is helpful in estimation, especially with nonlinear g.

Our estimators are the basis for new tests for covariate effects in Section 3. It is often of interest to investigate the functional form of the coefficients. In Section 4, we introduce estimators for parametric models which minimize a least squares criterion for the difference between the nonparametric estimate of $\beta_i(t)$ and the model. Their consistency and asymptotic normality are derived in the appendix. In Section 5, goodness-of-fit tests for the submodel are examined. The tests allow estimation of the parameters using methods other than those in Section 4 and are widely applicable. In Section 6, we discuss inferences for smooth functionals of the nonparametric and parametric estimates of the coefficients. These enable the construction of confidence regions for $E\{Y(t)|X(t), S(t) = 1\}$ which combine the two sets of estimates.

In recurrent event simulations, the methods perform well with realistic sample sizes. The efficiency of the parametric estimates appears competitive with those in Lin et al. (2001). The results are reported in Section 7. The practical utility of our procedures is illustrated in a reanalysis of graft versus host disease in a bone marrow transplant study (Pepe and Couper, 1997) in Section 8. Remarks conclude in Section 9.

2. Functional Estimating Equations

Within a time interval $[l, u]$, we continuously observe n independent and identically distributed copies of $\{Y(t), X(t), S(t) : R(t) = 1\}$, where Y is the response, X is a $p \times 1$ covariate vector, S is an implicit screening or stratifying variable, and R is the data availability indicator, which permits both missing response and missing covariates. The data is $\{Y_i(t), X_i(t), S_i(t) : R_i(t) = 1\}, i = 1, \ldots, n$. We assume that the data availability is independent of the response at each time t, conditioning on the covariates $X(t)$ and $S(t)$. In particular, we posit

$$E\{Y(t)|X(t), S(t) = 1\} = E\{Y(t)|X(t), S(t) = 1, R(t) = 1\}. \tag{2}$$

The assumption (2) is essentially saying that the missingness is noninformative.

The estimator for $\beta(t)$ in model (1) may be computed separately at each t. Define $\hat{\beta}(t)$ as the root of $U\{\beta(t), t\} = \sum_{i=1}^{n} A_i\{\beta(t), t\}$, where

$$A_i\{\beta(t), t\} = S_i(t)R_i(t)D_i^T\{\beta(t)\}V_i\{\beta(t), t\}\left[Y_i(t) - g^{-1}\{\beta^T(t)X_i(t)\}\right],$$

$D_i\{\beta(t)\} = \mathrm{d}[g^{-1}\{\beta^T(t)X_i(t)\}]/\,\mathrm{d}\{\beta(t)\}$ and $V_i\{\beta(t), t\}$ is a weight matrix, possibly random. The estimator potentially jumps at those M times where $\{Y_i(t), X_i(t), S_i(t), R_i(t)\}$ jumps. Let $j_1 < \ldots < j_M$ be the jump points. If $Y_i(t)$ and $X_i(t)$ are piecewise constant, then so also is the estimator and finding $\hat{\beta}$ involves solving U at the M points. In theory, when the processes vary between j_i, smoothing is not required. In practice, the equations are solved on a grid and the estimators are interpolated via smoothing.

The functional estimating equation U is an infinite dimensional analog of that in Liang and Zeger (1986). We adopt an independence working assumption across t. This avoids modeling the temporal correlations. Incorporating weights which account for such dependencies may improve the efficiency of the estimators. With longitudinal data, the response dimension is small and the specification is straightforward; see Prentice and Zhao (1991) and Zhao et al. (1992). Misspecifying the covariance does not bias the estimators. It is not obvious that this approach can be employed with high dimensional temporal processes. Nadeau and Lawless (1998) consider optimality in a class of linear estimating equations for parametric mean models which disregard information in associations over time.

Assume that Y, X, S, and R have finite jumps. Also, assume for each $t \in [l, u]$, $\mathrm{pr}\{R(t) = 1|X(t), S(t)\} > 0$ for all $\{X(t), S(t)\}$, that is, there is a positive probability of complete data. Under mild conditions (Liang and Zeger, 1986), for each t, $\hat{\beta}(t)$ is consistent for $\beta_0(t) = \{\beta_{10}(t), \ldots, \beta_{p0}(t)\}^T$, the true value of $\beta(t)$. Note that misspecification of model (1) at times other than t does not affect the validity of the estimates at t. Under the assumed model, for any $K < \infty$ points, $l < t_1 < \ldots, t_K < u$, $n^{1/2}[\{\hat{\beta}(t_1)^T, \ldots, \hat{\beta}(t_K)^T\}^T - \{\beta_0(t_1)^T, \ldots, \beta_0(t_K)^T\}^T]$ is asymptotically normal with covariance con-

sistently estimated by the "sandwich" estimator.

In the appendix, we show that the results are uniform in t. That is, $\hat{\beta}(t)$ converges uniformly to $\beta_0(t)$ for $t \in [l, u]$ and $n^{1/2}\{\hat{\beta}(t) - \beta_0(t)\}$ converges weakly to a tight Gaussian process with continuous sample paths at continuity points of $\beta_0(t)$; see Theorems 1 and 2. The covariance function $\Sigma(s,t) = \text{cov}[n^{1/2}\{\hat{\beta}(s) - \beta_0(s)\}, n^{1/2}\{\hat{\beta}(t) - \beta_0(t)\}] = \{H(s)^{-1}\}G(s,t)\{H(t)^{-1}\}^T$, where $H(s)$ and $G(s,t)$ are the asymptotic limits of

$$\hat{H}(s) = n^{-1}\sum_{i=1}^{n} S_i(t)R_i(t)D_i^T\{\hat{\beta}(t)\}V_i\{\hat{\beta}(t),t\}D_i\{\hat{\beta}(t)\} \text{ and } \hat{G}(s,t) = n^{-1}\sum_{i=1}^{n} A_i\{\hat{\beta}(s),s\}A_i\{\hat{\beta}(t),t\}^T.$$

Since the processes in U may be non-Markov, martingales are not applicable and empirical process theory is used to establish the results.

Pointwise confidence intervals for $\beta_0(t)$ may be constructed using the normal approximation and the variance estimate $\hat{\Sigma}(t,t) = \{\hat{H}(t)^{-1}\}\hat{G}(t,t)\{\hat{H}(t)^{-1}\}^T$. A $100(1-2\alpha)$ confidence interval at time t for $\beta_{i0}(t)$ is $\hat{\beta}_i(t) \pm n^{-1/2}z_\alpha\hat{\Sigma}_i(t,t)^{1/2}$, where z_α is the $(1-\alpha)$ percentile of the standard normal distribution and $\hat{\Sigma}_i(t,t)$ is the i-th diagonal element of $\hat{\Sigma}(t,t)$. Constructing confidence bands for $t \in [l,u]$ appears analytically intractable because the Gaussian process does not have a canonical representation. Instead, resampling may be employed, either bootstrapping the empirical data distribution and solving U repeatedly, or simulating directly from the process, as in Lin et al. (1994). Computational details and theoretical justification for the resampling are provided in the appendix; see Theorem 3.

3. Nonparametric Hypothesis Testing

We consider the null hypothesis $H_0 : C(t)\beta(t) = c(t)$, where at each t, $C(t)$ is an $r \times p$ contrast matrix and $c(t)$ is an $r \times 1$ vector of constants. This general framework allows global tests for multiple hypotheses. In the special case of testing the effect of the i-th covariate, one takes $C(t)$ to be a $1 \times p$ vector with a one in the i-th position and zeros elsewhere and $c(t) = 0$.

Three statistics are proposed for evaluating H_0. The first statistic is based on testing H_0 at $K < \infty$ points, $h_1, \ldots, h_K$. Let $\hat{\beta}^* = \{\hat{\beta}(h_1), \ldots, \hat{\beta}(h_K)\}$, $C^* = \text{diag}\{C(h_1), \ldots, C(h_K)\}$ and $c^* =$

$\{c(h_1)^T, \ldots, c(h_K)^T\}^T$. The statistic is

$$T_1 = (C^*\hat{\beta}^* - c^*)^T (C^*\hat{\Sigma}^* C^{*T})^{-1} (C^*\hat{\beta}^* - c^*),$$

where $\hat{\Sigma}^*$ is the estimated variance of $\hat{\beta}^*$ derived from $n^{-1}\hat{\Sigma}(s,t)$ in Section 2. The second statistic is based on the integrated difference $\Delta = \int_l^u \{C(t)\hat{\beta}(t) - c(t)\} W(t)\,\mathrm{d}t$, where W is a non-negative weight function, possibly random, with limit W^* as $n \to \infty$. Define $T_2 = \Delta^T \hat{\Sigma}_\Delta^{-1} \Delta$, where $\hat{\Sigma}_\Delta$ is the estimated covariance matrix of Δ,

$$\hat{\Sigma}_\Delta = n^{-2} \sum_{i=1}^{n} \left(\int_l^u C(s)\hat{H}(s)^{-1} A_i \left\{ \hat{\beta}(s), s \right\} W(s)\,\mathrm{d}s \right)^{\otimes 2},$$

and for a vector v, $v^{\otimes 2} = vv^T$. The third statistic is based on sup-norm distance,

$$T_3 = \sup_{t \in [l,u]} n |\{C(t)\hat{\beta}(t) - c(t)\}^T \{C(t)^T \hat{\Sigma}(t,t) C(t)\}^{-1} \{C(t)\hat{\beta}(t) - c(t)\}|.$$

Under H_0, the limiting distributions of T_1 and T_2 can be evaluated explicitly, with the p-values computed directly. Under mild conditions, T_1 is asymptotically χ^2_{rK} and T_2 is asymptotically χ^2_r, where χ^2_d denotes a chi-squared distribution with d degrees of freedom. When $r = 1$, the test statistic $T_2' = \hat{\Sigma}_\Delta^{-1/2} \Delta$ has standard normal distribution asymptotically. Similarly to most Kolmogorov-Smirnov type statistics, the distribution of T_3 is rather complex and must be approximated by resampling. Note that for T_1, the inferences rely on the pointwise results for $\hat{\beta}(t)$, while for T_2 and T_3, the stronger uniform convergence is needed.

For T_2, one should choose $W(t)$ to accentuate anticipated deviations from H_0. For example, when testing the effect of a single covariate, taking $W(t) > 0$ yields a test which is sensitive to "stochastic ordering" alternatives where $|C(t)\beta(t)| > c(t)$, $t \in [l, u]$. When the condition is violated, T_1 may have increased power, even with the increase in degrees of freedom. However, the choice of time points is somewhat arbitrary and may miss differences from the null at some t. The statistic T_3 is omnibus to all departures from H_0. A drawback is that such statistics are known to have low power because of a lack of specificity. In addition, it is computationally intensive compared to the other tests.

4. Parametric Estimation of $\beta(t)$

We consider the model $L^T(t)\beta(t) = f(\eta, t)$, where $L(t)$ is a $p \times 1$ contrast vector, f is a known function, continuously differentiable in η and t, and $\eta = (\eta_1, \ldots, \eta_g)^T$ is a $g \times 1$ vector of parameters. The set-up permits modeling differences in covariate effects, with $L(t)$ defining the contrasts of interest. One models $\beta_i(t)$ by letting $L(t)$ be 1 in the i-th position and 0 elsewhere. Polynomials with $f(\eta, t) = \sum_{j=1}^{g} \eta_j t^{j-1}$ are natural for time varying effects. A time independent coefficient occurs when $g = 1$ and $\beta_i(t) = \eta_1$.

Estimation of η minimizes the distance between $L(t)^T\hat{\beta}(t)$ and $f(\eta, t)$ in integrated squared error. Define $Q(\beta, \eta) = \int_l^u \{L^T(t)\beta(t) - f(\eta, t)\}^2 \tilde{W}(t)\,\mathrm{d}t$, where $\tilde{W}$ is a nonnegative function, possibly random, with limit $\tilde{W}^*$ as $n \to \infty$, and let $\hat{\eta} = \arg\min_\eta \{Q(\hat{\beta}, \eta)\}$. The theoretical properties of the estimator for the model for $\beta_i(t)$ do not assume the validity of the models for $\beta_j(t), j \neq i$. That is, the functional form of the coefficient for a particular covariate may be analyzed separately from other covariates. In the appendix, we show that if model (1) holds and $f(\eta, t)$ is specified correctly, then as $n \to \infty$, a unique $\hat{\eta}$ exists and is consistent for $\eta_0 = (\eta_{1,0}, \ldots, \eta_{g,0})^T$, the true value of η; see Theorem 4.

Observe that $\hat{\eta}$ is a solution to $\tilde{U}(\hat{\beta}, \eta) = 0$, where

$$\tilde{U}(\beta, \eta) = \partial\{Q(\beta, \eta)\}/\partial\eta = \int_l^u \dot{f}(\eta, t)^T \left\{L^T(t)\beta(t) - f(\eta, t)\right\} \tilde{W}(t)\,\mathrm{d}t,$$

and $\dot{f}(\eta, t) = d\{f(\eta, t)\}/d\eta$. A Taylor expansion of $\tilde{U}(\hat{\beta}, \hat{\eta})$ in $\hat{\eta}$ about η_0 gives

$$n^{1/2}(\hat{\eta} - \eta_0) = I(\beta_0, \eta_0)^{-1} \int_l^u \dot{f}(\eta_0, t)^T \left\{L^T(t)\hat{\beta}(t) - f(\eta_0, t)\right\} \tilde{W}^*(t)\,\mathrm{d}t + o_p(1), \tag{3}$$

where $I(\beta, \eta) = \int_l^u \dot{f}(\eta, t)\dot{f}^T(\eta, t)\tilde{W}^*(t)\,\mathrm{d}t$. Because of the uniform convergence of $\hat{\beta}$, under mild conditions, one may substitute the asymptotic equivalent for $L^T(t)\{\hat{\beta}(t) - \beta_0(t)\} = L^T(t)\hat{\beta}(t) - f(\eta_0, t)$ into (3), yielding $n^{-1/2}\sum_i \imath_i(\beta_0, \eta_0)$, where the influence function

$$\imath_i(\beta, \eta) = I(\beta, \eta)^{-1} \int_u^l \dot{f}(\eta, t) L^T(t) H(t)^{-1} A_i \left\{\beta(t), t\right\} \tilde{W}^*(t)\,\mathrm{d}t.$$

See the proof of Theorem 4 for details. A central limit theorem then gives that $n^{1/2}(\hat{\eta}-\eta_0)$ has limiting normal distribution with variance Γ which is consistently estimated by $\hat{\Gamma} = n^{-1}\sum_i \hat{\imath}_i(\hat{\beta},\hat{\eta})^{\otimes 2}$, where $\hat{\imath}_i$ is $\imath_i$ with H and $\tilde{W}^*$ replaced by $\hat{H}$ and $\tilde{W}$.

The weight $\tilde{W}$ may influence the variance of $\hat{\eta}$. For weighted least squares analyses of univariate data with heteroscedasticity, an optimal choice is the inverse of the variance. This suggests $\tilde{W}_1(t) = \{L^T(t)\hat{\Sigma}(t,t)L(t)\}^{-1}$, which emphasizes those t where $\hat{L}(t)^T\beta(t)$ is more precise. Simulations demonstrate efficiency gains relative to $\tilde{W}_2(t) = 1$.

When fitting $\beta_i(t) = \eta_1$, the estimator has a closed form

$$\hat{\eta}_1 = \int_l^u \hat{\beta}_i(t)\tilde{W}(t)\,\mathrm{d}t \left\{\int_l^u \tilde{W}(t)\,\mathrm{d}t\right\}^{-1},$$

which is a weighted average of the nonparametric estimator across time. Evidently, there is a correspondence between tests for $\beta_i(t) = 0$ from $\hat{\eta}_1$ and from the test statistic T_2 in Section 3. The estimator is closely related to the quantity Δ which was used to construct T_2, in that $\hat{\eta}_1$ is the root of $\Delta = 0$.

5. Goodness-of-fit Testing for $f(\eta, t)$

In exploratory analyses, it may be important to test the goodness-of-fit of models for $\beta_i(t)$. In some cases, estimates for η may be derived using procedures other than those in Section 4. For example, with recurrent event data, partial likelihood estimators are available for the Andersen and Gill (1982) model, where $g = \log$ in model (1). In general, we require that the estimator for η, $\tilde{\eta}$, is consistent and asymptotically normal. It is assumed that $n^{1/2}(\tilde{\eta}-\eta_0) = n^{-1/2}\sum_i \imath_i^* + o_p(1)$, where $\imath_i^*, i = 1,\ldots,n$, are mean zero and i.i.d., and that $\mathrm{var}\{n^{1/2}(\tilde{\eta}-\eta_0)\}$ may be consistently estimated by $n^{-1}\sum_i \hat{\imath}_i^*$, where $\hat{\imath}_i^*$ is the estimated influence function. Almost all estimators of practical interest have these properties. For $\hat{\eta}$, $\imath_i^* = \imath_i$ from Section 4.

The null hypothesis is that the model for $L^T(t)\beta(t)$ is correctly specified. That is, $H_0^* : L^T(t)\beta(t) = f(\eta, t)$. We construct tests by modifying $T_i, i = 1,2,3$, from Section 3. The idea is to use $J(t) =$

$L^T(t)\hat{\beta}(t) - f(\tilde{\eta}, t)$ in place of $C(t)\hat{\beta}(t) - c(t)$. A difficulty is that $f(\tilde{\eta}, t)$ is estimated, and not deterministic like $c(t)$ in H_0.

Using first order approximations, we can show that under H_0^*, $n^{1/2}J(t)$ is asymptotically equivalent to $n^{-1/2}\sum_i P_i(t)$, where

$$P_i(t) = L^T(t)H^{-1}(t)A_i\{\beta_0(t), t\} - \dot{f}(\eta_0, t)^T \imath_i^*.$$

Furthermore, since $\hat{\beta}(t)$ and $\tilde{\eta}$ are both tight, so is $n^{1/2}J(t)$ and it converges weakly to a Gaussian process with covariance function $\Omega(s,t)$ which is consistently estimated by $\hat{\Omega}(s,t) = n^{-1}\sum_i \hat{P}_i(s)\hat{P}_i(t)^T$, where $\hat{P}_i$ is P_i with H, β, η, and $\imath_i^*$ replaced by $\hat{H}, \hat{\beta}, \hat{\eta}$, and $\hat{\imath}_i^*$. Let $J^* = \{J(h_1), \ldots, J(h_Q)\}$. It follows that $n^{1/2}J^*$ has a Q-variate normal distribution with variance Ω^* which is consistently estimated by $\hat{\Omega}^* = n^{-1}\sum_i\{\hat{P}_i(h_1), \ldots, \hat{P}_i(h_Q)\}^{\otimes 2}$.

The first statistics is $T_1^* = \{n^{1/2}J^{*T}\}\{\hat{\Omega}^*\}^{-1}\{n^{1/2}J^*\}$. The second statistic $T_2^* = \hat{\Sigma}_{\Delta^*}^{-1/2}\Delta^*$ is based on the integrated difference $\Delta^* = \int_l^u J(t)W(t)\,\mathrm{d}t$, where $\hat{\Sigma}_{\Delta^*} = n^{-2}\sum_i\left\{\int_l^u \hat{P}_i(t)\tilde{W}(t)\,\mathrm{d}t\right\}^{\otimes 2}$ is the estimated variance of Δ^*. The third statistic is $T_3^* = \sup_{t\in[l,u]} n|J^2(t)\hat{\Omega}(t,t)^{-1}|$.

The distributions of T_1^* and T_2^* under H_0^* are analytically tractable: T_1^* is asymptotically χ_Q and T_2^* is asymptotically $N(0,1)$. The null limiting distribution of the sup-norm test depends on the covariance of P_i, which is rather complicated. The distribution may be approximated numerically by modifying the resampling technique described in Theorem 3.

6. Combining Estimates of $\beta(t)$

After assessing $\beta_j(t) = f_j(\eta_j, t)$, where $\eta_j = (\eta_{j,1}, \ldots, \eta_{j,g_j})^T$ with $g_j < \infty$ for $j = 1, \ldots, p$, it may be desirable to predict $E\{Y(t)|X(t), S(t) = 1\}$ for particular values of $X(t)$. If the model for β_j is correctly specified, then either the nonparametric procedures in Section 2 or parametric estimates of $f_j(\eta_j, t)$ may be used. For $j = 1, \ldots, p$, let $\tilde{\eta}_j$ denote a consistent and asymptotically normal estimator for η_j. Assume that $n^{1/2}(\tilde{\eta}_j - \eta_j^0) = n^{-1/2}\sum_i \imath_{ij}^* + o_p(1)$, where η_j^0 is the true value of η_j and for $j = 1, \ldots, p$, $\imath_{ij}^*$ are mean zero and i.i.d., with variance which may be consistently estimated

by $n^{-1}\sum_i \hat{\imath}^*_{ij}$, where $\hat{\imath}^*_{ij}$ is the estimated influence function. The estimator $\hat{\eta}$ in Section 4 has these properties.

Combining the estimators can be cast as the estimation of a smooth functional. Let $F^*(t)$ equal $F\{\beta, \eta_1, \ldots, \eta_p, t\}$ with $(\beta = \beta_0, \eta_j = \eta_j^0, j = 1, \ldots, p)$, where F is deterministic and depends only on $\beta(t)$ at t. Suppose $\partial F/\partial\beta = l_\beta$ and $\partial F/\partial\eta_j = l_{\eta_j}$, $j = 1, \ldots, p$, exist at $\beta = \beta_0$ and $\eta_j = \eta_j^0, j = 1, \ldots, p$, and are equicontinuous and uniformly bounded for $t \in [l, u]$. Since $\hat{\beta}$ is uniformly consistent for β_0 and $\tilde{\eta}_j$ is consistent for $\eta_j^0, j = 1, \ldots, p$, $\hat{F}^*(t) = F\{\hat{\beta}(t), \tilde{\eta}_1, \ldots, \tilde{\eta}_p, t\}$ converges uniformly to $F^*(t)$. If F is independent of $\beta(t)$, then estimation is based only on the parametric estimates. Conversely, if F is independent of $\eta^j, j = 1, \ldots, p$, then only $\hat{\beta}(t)$ is used.

Constructing confidence intervals and bands for F^* involves the limiting distribution of $\tilde{F}(t) = n^{1/2}\{\hat{F}^*(t) - F^*(t)\}$. Taylor expansions give that $\tilde{F}(t)$ and

$$l_\beta\{\beta_0(t), \eta_1^0, \ldots, \eta_p^0, t\}n^{1/2}\{\hat{\beta}(t) - \beta_0(t)\} + \sum_{j=1}^{p} l_{\eta_j}\{\beta_0(t), \eta_1^0, \ldots, \eta_p^0, t\}n^{1/2}(\tilde{\eta}_j - \eta_j^0)$$

are asymptotically equivalent. Using the influence functions for $\hat{\beta}$ and $\tilde{\eta}_j$, we can establish that $\tilde{F}(t)$ converges weakly to a tight Gaussian process with $\mathrm{cov}\{\tilde{F}(s), \tilde{F}(t)\} = \Phi(s, t)$ which is consistently estimated by $\hat{\Phi}(s, t) = n^{-1}\sum_i \hat{\phi}_i(s)\hat{\phi}_i(t)$, where

$$\hat{\phi}_i(s) = l_\beta\{\hat{\beta}(t), \tilde{\eta}_1, \ldots, \tilde{\eta}_p, t\}\hat{H}^{-1}(t)A_i\{\hat{\beta}(t), t\} + \sum_{j=1}^{p} l_{\eta_j}\{\hat{\beta}(t), \tilde{\eta}_1, \ldots, \tilde{\eta}_p, t\}\hat{\imath}^*_{ij}.$$

Employing the normal approximation, a $100(1 - 2\alpha)$ confidence interval for $F^*(t)$ is given by $\hat{F}^*(t) \pm n^{-1/2}z_\alpha\hat{\Phi}(t, t)^{-1/2}$. Simultaneous confidence bands may be computed using the influence function simulation technique in the appendix.

7. Recurrent Event Simulations

To investigate the performance of the estimators and test statistics, we conducted numerical studies with data generated under the recurrent event set-up in Lin et al. (2001). Let $N(t)$ be the number

of events by t and let Z be a binary covariate, equal to either 0 or 1 with probability 0.5. The proportional means model is

$$E\{N(t)|\psi, Z) = \psi f\left[\exp\left\{\alpha(t) + \beta(t)Z\right\}\right], \tag{4}$$

where ψ is an independent gamma random variable with mean 1 and variance σ^2 and $\alpha(t)$ is the log of baseline intensity function. Taking $f(x) = x$ gives $E\{N(t)|Z\} = \exp[\alpha(t) + \beta(t)Z]$, the proportional means model. We set $\alpha(t) = \log(t)$, $\beta(t) = 0.5$, and σ^2 =0, 0.25, 0.5 or 1. The censoring time was independently generated from a $[c - c_0, c]$ distribution, with $R(t) = I(C > t)$. In the following, $c = 3$ and $c_0 = 0$, 1 or 3, yielding 4.0, 3.3 and 2.0 observed events on the average per subject, respectively. For each combination of σ^2 and c_0, we generate 1000 datasets with $n = 100$.

The functional estimating equation $U\{\beta(t), t\}$ provides estimates of $\alpha(t)$ and $\beta(t)$ at each t in $(0, c)$. Note that times in the right tail may have a small number of observations that are uncensored, that is, with $R(t) = 1$. This may lead to instability in $\hat{\eta}$ and in the test statistics. The lower and upper endpoints, l and u, in T_2 and $Q\{\hat{\beta}(t), \eta\}$ must be chosen carefully. We take $l = 0.25$ and take $u = 3$ when $c_0 = 0$ and equal to the 75th and 90th percentile of the observed censoring distribution when c_0 =1 or 3. We use two weights in Q: the inverse variance function $\tilde{W}_1(t)$ and $\tilde{W}_2(t) = 1$.

[Table 1 about here.]

The results are summarized in Table 1 and include the average bias, the empirical variance, the average model based variance, and the coverage probabilities of the 0.95 confidence intervals for $\beta(t) = \eta$. The estimator $\hat{\eta}$ is virtually unbiased and the empirical variances generally agree with the model based variances. As c_0 increases, the model based variance may underestimate the empirical variance slightly and the resulting coverage probabilities are somewhat liberal. The method used to determine the upper endpoint u does not have much effect on the results. Using $\tilde{W}_1$ is more efficient than $\tilde{W}_2$.

Also given in Table 1 is the ratio of the empirical variance of $\hat{\beta}_F$, the estimator recommended by Lin et al. (2001), to that of $\hat{\eta}$. The weighted and unweighted estimators are both rather competitive and are computationally simpler than $\hat{\beta}_F$. When using $\tilde{W}_1$, the relative efficiency is close to one with light censoring and above 0.80 in almost all cases.

Next, the data was generated with different f in (4). As in Lin et al. (2001), $f_1(x) = x, f_2(x) = (1+x)^2 - 1$, and $f_3(x) = \log(1+x)$. In general, $E\{N(t)|Z\} = g^{-1}\{\alpha(t) + \beta(t)Z\}$, where $g = f \circ \log$. As before, $\alpha(t) = \log(t)$ and $\beta(t) = 0.5$. The censoring parameters c_0 and c were chosen so that there were on average 3 observed events per subject and $\sigma^2 = 0.25$. We simulated 1000 data sets with $n = 100$ and estimated $\hat{\eta}$ using $\tilde{W}_1$. To test $H_0 : \beta(t) = 0.5$, T_1 was calculated with one, two, and four time points chosen to partition the segment from l to u into two, three, and five equal length pieces, respectively. For T_2, $W(t) = 1$. The bias, and empirical and model based variances of $\hat{\eta}$, and the relative efficiency of $\hat{\eta}$ to $\hat{\beta}_F$, are reported in Table 2. The bias is small and the empirical and model based variances tend to agree. The tests reject at close to the nominal rate, although T_1 is somewhat anti-conservative as the number of points increases, due to unstable estimation of $\hat{\Sigma}^*$. With non-proportional means models, f_2 and f_3, the estimator from Q may have smaller variance than that in Lin et al. (2001).

[Table 2 about here.]

In other simulations, a time-varying effect was included in (4). We let $\beta(t) = \eta_1 + \eta_2 t$, with $\eta_1 = 0$ and $\eta_2 = 1/c$. Note that $c^{-1}\int_0^c \beta(t)\,\mathrm{d}t = 0.5$, that is, the average effect of Z is the same as in the model with the time-independent covariate. In these analyses, we let u be the 75th percentile of the empirical censoring distribution and set $l = c - u$. The results for $\hat{\eta}_2$ with $\tilde{W}_1$ and the test statistics for H_0 are in Table 3. The estimator behaves well, as with time-independent coefficients. The one point chi-square test and the integral test are insensitive to deviations of $\beta(t)$ from 0.5. This happens because $\beta(c/2) = 0$ and $\int_l^u \beta(t)\,\mathrm{d}t = 0.5$ in each dataset. On the other hand, T_1 with two and four

points appears to have substantial power.

[Table 3 about here.]

8. Prevalence Data

We reanalyze data on chronic graft-versus-host disease (CGVHD) in bone marrow transplantation from Pepe and Couper (1997). There are 147 subjects, each randomly assigned to methotrexate plus cyclosporine (MC) or prednisone plus MC (PMC) as prophylaxis for acute graft versus host disease. Patients' age and sex were recorded, along with other covariates. Our focus is the effects of prophylaxis, sex and age, on the prevalence of CGVHD among those alive and relapse free.

Let $Y_i(t)$ be 1 if patient i has CGVHD at t and 0 if not. Let $S_i(t)$ be 1 if patient i is alive and relapse-free at t and 0 if not. Let $R_i(t)$ be 0 if patient i has been lost to follow-up at t and 1 if not. As in Pepe and Couper (1997), a logistic prevalence model is assumed:

$$\text{logit}[E\{Y_i(t)|X_i, S_i(t) = 1\}] = \alpha(t) + \beta^T(t)X_i. \tag{5}$$

The model is restricted to $t \in [l, u]$, where $l = 83$ days and $u = 974$ days, to ensure parameter estimability. In the sequel, these values for l and u are used in Q, T_2 and T_2^*.

We first fit the prevalence function with treatment as the only covariate (1 for PMC, 0 for MC). The statistic $T_1 = 10.23$ using $\hat{\beta}(t)$ at two time points equal to the 15th and 85th percentiles of the jump points of $\{Y_i(t), S_i(t), R_i(t), i = 1, \ldots, 147\}$ and $T_2 = -2.35$ with $W = \tilde{W}_1$, yielding p-values 0.006 and 0.019, respectively. Thus, H_0 is rejected. Next, the parametric model $\beta(t) = \eta$ was fit with $\tilde{W} = \tilde{W}_1$, as described in Section 4. The estimate $\hat{\eta} = 0.91$, with standard error 0.38. The estimate of η in Pepe and Couper (1997) was 0.92, with standard error 0.39. The results are in close agreement.

[Figure 1 about here.]

The estimated prevalences for MC ($X_i = 0$) and PMC ($X_i = 1$) using $(\hat{\alpha}, \hat{\beta})$ are shown in Figure 1. These curves are equivalent to the nonparametric estimators in Section 3.1 of Pepe et al. (1991),

that is, $\sum_i Y_i(t)S_i(t)R_i(t)I\{X_i(t) = \text{trt}\}[\sum_i S_i(t)R_i(t)I\{X_i(t) = \text{trt}\}]^{-1}$ for trt $= 0, 1$, where I is the indicator function. Also displayed are estimates from $(\hat{\alpha}, \hat{\eta})$, along with 0.95 pointwise confidence intervals. There is a perfect match between the nonparametric and parametric estimates on MC because $\hat{\eta}$ vanishes when $X_i = 0$. In the other arm, the two curves are similar until 800 days, but diverge at later times.

[Figure 2 about here.]

We now include treatment, patient gender (1 for males, 0 for females) and patient age in the model. The nonparametric estimates of the coefficients are pictured in the Figure 2. A visual inspection suggests linear models for treatment and gender, and a constant model for age. The corresponding parametric fits are overlaid, along with 0.95 pointwise confidence intervals and 0.95 confidence bands from the nonparametric estimate. The bands are noticeably wider than the intervals (≈ 1.7 times). This is because the correlation between $\hat{\beta}(s)$ and $\hat{\beta}(t)$ is weak, decreasing rapidly as $|s - t|$ increases. The differences between the two sets of curves generally decrease as the correlation increases, with perfect agreement when the correlation is one for all s, t.

[Table 4 about here.]

Table 4 summarizes the hypothesis testing for $\beta_i(t) = 0$ and the goodness-of-fit testing for constant and linear submodels for each covariate effect. Test statistic T_1 is based on two time points s_1 and s_2 equal to the 15th and 85th percentiles of the jumps. Test statistic T_2' is based on $\tilde{W}(t) = [n\text{var}\{\hat{\beta}_i(t)\}]^{-1}$. Prednisone and maleness decrease prevalence, with the benefit appearing to increase over time. The tests of age suggest that older patients may be at greater risk. Below, we formally examine the time-dependence of the regression parameters.

For each covariate, we begin by fitting a constant model. If the constant model fits poorly, then the linear model is fitted and tested for goodness-of-fit. The statistic T_1^* is based on s_1 and s_2 equal to the

15th and 85th percentiles of the jumps. When lack-of-fit is present, the integration in T_2^* may cancel deviations from early and late time points when $\beta_i(t)$ is not "stochastically ordered" relative to the fitted model. For comparison, we compute T_{2a}^* with $W(t) = 1$, T_{2b}^* with $W(t) = I\{t < (u-l)/2\}$ and T_{2c}^* with $W(t) = I\{t > (l-u)/2\}$, where $I(\cdot)$ is the indicator function. For treatment and patient gender, concluding that the linear model fits better than the constant seems justified by the graphical displays and the numerical tests, while for patient age the time-independent coefficient may be adequate. The "best fitting" modela are $\hat{\beta}_{\text{trt}} = -0.4372 - 0.0013\ t$ for treatment, $\hat{\beta}_{\text{gender}} = 0.0336 - 0.0020\ t$ for patient gender, and $\hat{\beta}_{\text{age}} = 0.0366$ for patient age.

9. Remarks

The functional regression model is natural for complex event history data when the effects of covariates on the marginal distribution of the response is the primary interest. A limitation is that the estimating equation $U\{\beta(t), t\}$ is only applicable when the observation windows are known. This is not the case with right censoring, unless the censoring time is available on all subjects. When this occurs, there is a recurrent event structure and the number of events $Y(t)$ is ≤ 1.

Because the model (1) refers to means and not intensities, and does not involve a Markov assumption, the parameters are interpreted conditionally on covariates at t, and not all $s \leq t$. The distinction is important with time-dependent $X(t)$ when $S(t) = 1$, all t. In particular, with survival times, the model for the failure probability at t defines a hazard model when $X(t) = X$, but may not when $X(t)$ varies over time. That is, the failure probability may not denote the survivor function, as it would if the conditioning set included $\{X(s), s \leq t\}$. Interestingly, with time-independent covariates, the Aalen (1980) model obtains with $g = \log$.

Appendix A

Uniform Consistency and Weak Convergence of $\hat{\beta}(t)$

The results use the empirical process methods in van der Vaart and Wellner (1996) (hereafter VW). Regularity conditions are now stated. The coefficient $\beta(t)$ is right continuous with left hand limits but otherwise unspecified (hereafter cadlag). Restrict $h = g^{-1}$ and $\dot{h}(u) = \partial h(u)/\partial u$ to be Lipschitz continuous and bounded on compacts. Thus, $D_i\{\beta(t)\} = X_i(t)\dot{h}\{\beta^T(t)X_i(t)\}$. Assume that the data are cadlag, that $S_i(t)$, $R_i(t)$, and $X_i(t)$ have total variation on $[l, u]$ bounded by some $c < \infty$, almost surely, and that the total variation $\tilde{Y}_i$ of $Y_i(t)$ on $[l, u]$ has bounded second moment. We require that $\inf_{t\in[l,u]} \text{eigmin } E\{S_1(t)R_1(t)X_1(t)X_1^T(t)\} > 0$, where eigmin is the minimum eigenvalue of a matrix. Assume also that, for all bounded $B \subset \mathcal{R}^p$, the class of random functions $\{V_i(b, t) : b \in B, t \in [l, u)\}$ is bounded above and below by positive constants, is bounded in uniform entropy integral (hereafter BUEI) with bounded envelope, and is pointwise measurable (hereafter PM); see VW for definitions of BUEI and PM. For example, this is satisfied if $V_i(b, t) = V\{b^T X_i(t)\}$, where V is Lipschitz continuous.

Theorem 1. *Assume that model (1) holds with true parameter $\{\beta_0(t) : t \in [l, u]\}$ with $\sup_{t\in[l,u]} |\beta_0(t)| < \infty$. Let $\{\hat{\beta}(t) : t \in [l, u]\}$ be the smallest (in uniform norm) root of $U\{\beta(t), t\} = 0 : t \in [l, u]$. Then such a root exists for all n large enough, and $\sup_{t\in[l,u]} |\hat{\beta}(t) - \beta_0(t)| \to 0$, almost surely.*

Proof. Define $C(\gamma, \beta, t) = S(t)R(t)D\{\gamma(t)\}^T V\{\gamma(t), t\}[Y(t) - h\{\beta^T(t)X(t)\}]$, where the parameters $\gamma, \beta \in \{\ell_c^\infty([l, u])\}^p$ and $\ell_c^\infty(H)$ is the collection of bounded real functions on the set H with absolute value $\leq c$ (c is omitted if $c = \infty$). We first show that the class of functions $\mathcal{G} = \{C\{\gamma, \beta, t\} : \gamma, \beta \in \{\ell_c^\infty([l, u])\}^p, t \in [l, u]\}$ is BUEI with square-integrable envelope and PM for each $c < \infty$. This implies that $\mathcal{G}$ is Donsker by theorem 2.5.2 of VW and hence Glivenko-Cantelli. The result follows by the equivalence of the classes $\{\beta^T(t)X(t) : \beta \in \{\ell_c^\infty([l, u])\}^p, t \in [l, u]\}$ and $\{b^T X(t) : b \in [-c, c]^p, t \in [l, u]\}$, since cadlag processes bounded in total variation are Vapnic-Červonencis classes (and thus BUEI) and are PM, since Lipschitz continuous functions of BUEI and PM classes are also BUEI and

PM, and since both sums and products of BUEI and PM classes are also BUEI and PM. Therefore, for each $c < \infty$ and all $\tilde{\beta} \in \{\ell_c^\infty([l,u])\}^p$, $n^{-1}U\{\tilde{\beta}(t),t\}$ equals

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^n \Big(C_i(\tilde{\beta},\beta_0,t) + S_i(t)R_i(t)D_i^T\{\tilde{\beta}(t)\}V_i\{\tilde{\beta}(t),t\}\left[h\{\tilde{\beta}^T(t)X_i(t)\} - h\{\beta_0^T(t)X_i(t)\}\right]\Big)\\
=&n^{-1}\sum_{i=1}^n S_i(t)R_i(t)X_i(t)X_i^T(t)\dot{h}\{\check{\beta}^T(t)X_i(t)\}\dot{h}\{\tilde{\beta}^T(t)X_i(t)\}V_i\{\tilde{\beta}(t),t\}\left\{\tilde{\beta}(t)-\beta_0(t)\right\} - \epsilon_n(t), \quad \text{(A.1)}
\end{aligned}
$$

where $\check{\beta}$ is on the line segment between $\tilde{\beta}$ and β_0 and $\sup_{t\in[l,u]}|\epsilon_n(t)| \to 0$ almost surely, uniformly over $\tilde{\beta} \in \{\ell_c^\infty([l,u])\}^p$, since the terms involved have mean zero and are sums of Glivenko-Cantelli classes, hence are Glivenko-Cantelli. Since the eigenvalues of $E\{S_1(t)R_1(t)X_1(t)X_1^T(t)\}$ are bounded below, the above results imply $U\{\tilde{\beta}(t),t\} = 0$ has a uniformly bounded solution $\hat{\beta}(t)$ for all $t \in [l,u]$, for n large enough. Equation (A.1) then gives that $0 = \|n^{-1}U\{\hat{\beta}(t),t\}\| \geq k\|\hat{\beta}(t) - \beta_0(t)\| - \epsilon_n^*(t)$, where $k > 0$ does not depend on t, $\|\cdot\|$ is the Euclidean norm, and where $\sup_{t\in[l,u]}|\epsilon_n^*(t)| \to 0$ almost surely. Uniform consistency follows. □

Theorem 2. *Under the conditions of theorem 1, $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically linear with influence function $\psi_i(t) = \{H(t)\}^{-1}A_i\{\beta_0(t),t\}$, where $H(t) = E\left[S_1(t)R_1(t)D_1\{\beta_0(t)\}V_1\{\beta_0(t),t\}D_1^T\{\beta_0(t)\}\right]$, and converges weakly in $\{\ell^\infty([l,u])\}^p$ to a tight, mean zero Gaussian process $\mathcal{Z}(t)$ with covariance $E\{\mathcal{Z}(s)\mathcal{Z}^T(t)\} = E\{\psi_1(s)\psi_1^T(t)\}$.*

Proof. By theorem 1, for all n large enough,

$$
\begin{aligned}
0 &= n^{-1/2}U\{\hat{\beta}(t),t\}\\
&= n^{-1/2}U\{\beta_0(t),t\} + n^{-1/2}\left[U\{\hat{\beta}(t),t\} - U\{\beta_0(t),t\}\right]\\
&= n^{-1/2}U\{\beta_0(t),t\} + n^{-1/2}\sum_{i=1}^n\left\{C_i(\hat{\beta},\beta_0,t) - C_i(\beta_0,\beta_0,t)\right\}\\
&\quad + n^{-1/2}\sum_{i=1}^n S_i(t)R_i(t)D_i^T\{\hat{\beta}(t)\}V_i\{\hat{\beta}(t),t\}\left[h\{\hat{\beta}^T(t)X_i(t)\} - h\{\beta_0^T(t)X_i(t)\}\right]\\
&= U_0(t) + U_1(t) + U_2(t).
\end{aligned}
$$

Since $\mathcal{G}$ is Donsker and sums of Donsker classes are Donsker, and since $\sup_{t\in[l,u]} E\{C_1(\hat{\beta},\beta_0,t) - C_1(\beta_0,\beta_0,t)\}^2 \to 0$ in probability, $\sup_{t\in[l,u]} |U_1(t)| \to 0$ in probability. By previous arguments, $U_2(t) = \{H(t) + \epsilon_n^{**}(t)\}\sqrt{n}\left\{\hat{\beta}(t) - \beta_0(t)\right\}$, where $\sup_{t\in[l,u]} |\epsilon_n^{**}(t)| \to 0$. The implication is $\sqrt{n}\{\hat{\beta}(t) - \beta_0(t)\} = -\{H(t)\}^{-1}U_0(t) + \eta_n^*(t)$, where $\sup_{t\in[l,u]} |\eta_n^*(t)| \to 0$, giving asymptotic linearity. Weak convergence follows since $\{A_i\{\beta_0(t),t\} : t \in [l,u]\}$ is a subclass of the Donsker class $\mathcal{G}$. □

The next theorem establishes the consistency of $\hat{\Sigma}(s,t)$ and the validity of simultaneous confidence bands of the form

$$\hat{\beta}_j(t) \pm n^{-1/2}k_j(\alpha)\hat{M}_j(t), \tag{A.2}$$

$j = 1,\ldots p$, where the vector $\hat{M}(t)$ is the component-wise square root of the diagonal of $\hat{\Sigma}(t,t)$, $k_j(\alpha)$ is the $1-\alpha$ quantile of the empirical distribution of the j-th component of

$$K_n = n^{-1/2}\sum_{i=1}^{n} Z_i \mathrm{diag}\left[\{\hat{M}(t)\}^{-1}\right]\{\hat{H}(t)\}^{-1}A_i\{\hat{\beta}(t),t\},$$

from repeatedly sampling $Z_1,\ldots,Z_n$, where $Z_i,\ldots,Z_n$ are i.i.d. $N(0,1)$.

Theorem 3. *Under the conditions of theorem 1, $\hat{\Sigma}(s,t)$ is uniformly consistent for $\Sigma(s,t)$, over all $s,t \in [l,u]$, almost surely. If, in addition, $\inf_{t\in[l,u]} \mathrm{eigmin}\ \Sigma(t,t) > 0$, then the $1-\alpha$ simultaneous confidence bands given by (A.2) are asymptotically valid.*

Proof. Choose c so that $\sup_{t\in[l,u]} |\beta_0(t)| < c$, and let $B = \{\ell_c^\infty([l,u])\}^p$. Then $\hat{H}(t)$ is the average of i.i.d. processes indexed by $\beta \in B$ and $t \in [l,u]$ at $\beta = \hat{\beta}$. Since $\hat{\beta} \in B$ for all n large enough and the i.i.d. processes are Glivenko-Cantelli and sufficiently smooth, $\sup_{t\in[l,u]} \left|\hat{H}(t) - H(t)\right| \to 0$ in probability. Since $\mathcal{F} = \{A_i\{\beta(t),t\} : \beta \in B, t \in [l,u]\}$ is Donsker with square-integrable envelope, the class of all pairwise products in $\mathcal{F}$ is Glivenko-Cantelli. This, coupled with the properties of the i.i.d. processes, gives the uniform consistency of $\hat{G}$ and thus $\hat{\Sigma}$.

To establish validity of the confidence bands, we show that $\hat{M}(t)$ is uniformly consistent for $M(t)$, the component-wise square root of $\mathrm{diag}\{\Sigma(t,t)\}$, and that, with probability 1, the conditional law

of K_n, given the data, converges weakly to a tight, mean zero Gaussian process with covariance $\text{diag}\{M^{-1}(s)\}\Sigma(s,t)\text{diag}\{M^{-1}(t)\}$. The consistency of $\hat{M}$ follows from the consistency of $\hat{\Sigma}$. The convergence of the conditional law of K_n follows from the convergence of the conditional law of $L_n(\beta) = n^{-1/2}\sum_{i=1}^n Z_i A_i\{\beta(t), t\}$ at $\beta = \hat{\beta}$ combined with Slutsky's lemma for empirical processes. Since $\mathcal{F}$ is Donsker with square-integrable envelope, the almost sure convergence of the conditional law for $\{L_n(\beta) : \beta \in B\}$ follows by Theorem 2.9.7 in VW. Since the processes are sufficiently smooth, the conditional law of $L_n(\hat{\beta})$ converges to that of $L_n(\beta_0)$, and the results follow. □

Appendix B

Consistency and Asymptotic Normality of $\hat{\eta}$

We first state some additional regularity conditions. Assume that the model $L^T(t)\beta_0(t) = f(\eta, t)$ is identified. That is, there is a unique $\eta_0 \in \mathcal{R}^k$ satisfying the equality for all $t \in [l, u]$. Then, η_0 is a unique solution of $Q^*(\eta) = \int_l^u \left\{L^T(t)\beta_0(t) - f(\eta, t)\right\}^2 \tilde{W}^*(t)\,\mathrm{d}t = 0$. Assume also that $\liminf_{n\to\infty} Q^*(\eta_n) > 0$ for any sequence with $\|\eta_n\| \to \infty$, and that $f(\eta, t)$ is two times differentiable, such that $\int_l^u \dot{f}(\eta_0, t)\dot{f}^T(\eta_0, t)\tilde{W}^*(t)\,\mathrm{d}t$ is positive definite and $\int_l^u \left|\ddot{f}(\eta, t)\right|^2 \tilde{W}^*(t)\,\mathrm{d}t < \infty$ for all finite η, where $\ddot{f} = \partial^2 f(u)/\partial u^2$. Let $\hat{\eta}$ be the smallest of the minimizers in $\text{argmin}_\eta\, Q(\hat{\beta}, \eta)$ (if it is not unique).

Theorem 4. *Under the conditions of Theorem 1, and the stated regularity conditions, $\hat{\eta} \xrightarrow{p} \eta_0$; $\sqrt{n}(\hat{\eta} - \eta_0)$ is asymptotically linear, with influence function $\imath_i(\beta_0, \eta_0)$, and asymptotically mean zero normal with variance $\Gamma = E\left\{\imath_1(\beta_0, \eta_0)\imath_1(\beta_0, \eta_0)^T\right\}$, which is consistently estimated by $\hat{\Gamma}$.*

Proof. Consistency follows from Theorem 1 and the regularity conditions on $f(\eta, t)$. To see this, note that $\hat{\eta}$ is bounded in probability, otherwise there would exist a sequence (η_n) with $\eta_n \to \infty$ and $Q^*(\eta_n) \to 0$, which is impossible. Since all convergent subsequences of $\hat{\eta}$ satisfy $Q^*(\hat{\eta}) \to 0$, consistency follows. Asymptotic linearity and weak convergence follow from the Taylor expansion

given in (3), Theorem 2 and the fact that the class $\imath_i(\beta_0, \eta_0)$ is square-integrable since $A_i\{\beta_0(t), t\}$ has square-integrable envelope. Consistency of $\hat{\Gamma}$ follows from smoothness of $\dot{f}(\eta, t)$ and Theorem 3. $\square$

References

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. In Klonechi, W. and Rosinski, J., editors, *Lecture Notes in Statistics-2: Mathematical Statistics and Probability Theory*, pages 1–25.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1995). *Statistical models based on counting processes*. Springer-Verlag Inc.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study (Com: p1121-1124). *The Annals of Statistics* **10**, 1100–1120.

Fahrmeir, L. and Klinger, A. (1998). A nonparametric multiplicative hazard model for event history analysis. *Biometrika* **85**, 581–592.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Series B, Methodological* **62**, 303–322.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 757–796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–822.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association* **86**, 114–129.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96**, 103–126.

Lin, D. Y. (2000). Proportional means regression for censored medical costs. *Biometrics* **56**, 775–778.

Lin, D. Y., Fleming, T. R. and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81.

Lin, D. Y., Wei, L. J. and Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association* **96**, 620–628.

Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika* **86**, 691–702.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–514.

Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications* **39**, 153–180.

Nadeau, C. and Lawless, J. F. (1998). Inference for means and covariances of point processes through estimating functions. *Biometrika* **85**, 893–906.

Pepe, M. S. and Couper, D. (1997). Modeling partly conditional means with longitudinal data. *Journal of the American Statistical Association* **92**, 991–998.

Pepe, M. S., Heagerty, P. and Whitaker, R. (1999). Prediction using partly conditional time-varying coefficients regression models. *Biometrics* **55**, 944–950.

Pepe, M. S., Longton, G. and Thornquist, M. (1991). A qualifier Q for the survival function to describe the prevalence of a transient condition. *Statistics in Medicine* **10**, 413–421.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–839.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis (Disc: p561-572).

Journal of the Royal Statistical Society, Series B, Methodological **53**, 539–561.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics (ISBN 0387946403).* Springer-Verlag Inc.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association* **93**, 1388–1402.

Zhao, L. P., Prentice, R. L. and Self, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *Journal of the Royal Statistical Society, Series B, Methodological* **54**, 805–811.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *The Annals of Statistics* **18**, 329–353.

Table 1

Simulation results for $\hat{\eta}$ with $\tilde{W}_1$ and $\tilde{W}_2$ under $f(x) = x$ and $\beta(t) = 0.5$.

			Bias		ModVar		EmpVar		Cov95			
c_0	σ^2	u	$\tilde{W}_1$	$\tilde{W}_2$	$\tilde{W}_1$	$\tilde{W}_2$	$\tilde{W}_1$	$\tilde{W}_2$	$\tilde{W}_1$	$\tilde{W}_2$	RE_1	RE_2
0.0	0.00		0.001	0.005	0.014	0.018	0.014	0.019	0.947	0.949	0.90	0.68
	0.25		−0.007	−0.004	0.024	0.028	0.024	0.028	0.953	0.943	0.96	0.84
	0.50		−0.004	−0.002	0.034	0.037	0.036	0.040	0.947	0.941	0.90	0.82
	1.00		−0.004	−0.004	0.054	0.056	0.055	0.058	0.948	0.949	1.00	0.95
1.0	0.00	0.75	−0.003	0.002	0.017	0.020	0.017	0.021	0.952	0.954	0.93	0.76
		0.90	−0.002	0.002	0.017	0.019	0.016	0.019	0.952	0.951	0.99	0.85
1.0	0.25	0.75	−0.005	−0.002	0.027	0.030	0.027	0.029	0.951	0.946	0.98	0.90
		0.90	−0.003	−0.002	0.027	0.029	0.029	0.032	0.944	0.941	0.90	0.82
1.0	0.50	0.75	0.006	0.008	0.037	0.039	0.039	0.041	0.937	0.939	0.92	0.87
		0.90	−0.008	−0.004	0.037	0.040	0.041	0.044	0.932	0.928	0.88	0.82
1.0	1.00	0.75	0.008	0.011	0.058	0.060	0.063	0.066	0.929	0.930	0.95	0.91
		0.90	0.010	0.012	0.057	0.060	0.061	0.065	0.933	0.933	0.98	0.91
3.0	0.00	0.75	0.001	0.004	0.032	0.034	0.032	0.034	0.947	0.947	0.85	0.82
		0.90	−0.001	0.006	0.030	0.031	0.033	0.034	0.940	0.927	0.84	0.81
3.0	0.25	0.75	−0.005	0.000	0.046	0.048	0.055	0.057	0.932	0.934	0.77	0.73
		0.90	0.012	0.019	0.046	0.049	0.052	0.054	0.931	0.927	0.81	0.78
3.0	0.50	0.75	0.009	0.012	0.061	0.063	0.063	0.065	0.926	0.926	0.91	0.88
		0.90	0.015	0.020	0.061	0.065	0.069	0.073	0.911	0.921	0.83	0.79
3.0	1.00	0.75	−0.002	0.001	0.088	0.090	0.112	0.113	0.910	0.912	0.81	0.79
		0.90	−0.005	0.000	0.089	0.098	0.104	0.114	0.918	0.920	0.86	0.79

Bias, average bias; ModVar, average $\hat{\Gamma}$; EmpVar, empirical variance; Cov95, empirical coverage probabilities; RE_i, ratio of empirical variance of $\hat{\beta}_F$ to that of $\hat{\eta}$ with $\tilde{W}_i$

Table 2

Simulation results for $\hat{\eta}$ with $\tilde{W}_1$ and T_1 and T_2 with various $f(x)$ and $\beta(t) = 0.5$

$f(x)$	c_0	c	u	Bias	ModVar	EmpVar	RE_1	RR_{11}	RR_{12}	RR_{14}	RR_2
f_1	1.0	3.0	0.75	−0.005	0.027	0.027	0.98	0.053	0.055	0.059	0.053
	1.0	3.0	0.90	−0.003	0.027	0.029	0.90	0.055	0.060	0.073	0.060
f_2	0.3	0.9	0.75	0.011	0.019	0.020	1.03	0.057	0.057	0.067	0.050
	0.3	0.9	0.90	0.001	0.019	0.019	1.06	0.058	0.067	0.082	0.063
f_3	5.0	15.0	0.75	−0.001	0.145	0.147	0.88	0.052	0.050	0.075	0.057
	5.0	15.0	0.90	0.002	0.147	0.155	0.84	0.057	0.060	0.070	0.063

RR_{1j}, rejection rate for T_1, j time points; RR_2, rejection rate for T_2

Table 3

Simulation results for $\hat{\eta}_2$ with $\tilde{W}_1$ and T_1 and T_2 with various $f(x)$ and $\beta(t) = tc^{-1}$.

$f(x)$	c_0	c	Bias	ModVar	EmpVar	Cov95	RR_{11}	RR_{12}	RR_{14}	RR_2
f_1	1.0	3.0	−0.001	0.014	0.013	0.953	0.063	0.586	0.621	0.068
f_2	0.3	0.9	−0.005	0.172	0.144	0.927	0.065	0.583	0.692	0.071
f_3	5.0	15.0	0.004	0.002	0.002	0.946	0.049	0.293	0.306	0.050

Table 4

Hypothesis testing and goodness-of-fit testing for covariate effects in the C-GVHD prevalence analysis

	treatment		patient gender		patient age	
Test	Stat	p-value	Stat	p-value	Stat	p-value
Hypothesis testing of $H_0 : \beta_i(t) = 0$						
T_1	9.899	0.007	4.683	0.096	4.502	0.105
T_2'	−2.431	0.015	−2.005	0.045	1.970	0.049
Goodness-of-fit testing of constant submodel, $H_0^* : \beta_i(t) = \eta_1$						
T_1^*	7.208	0.027	2.471	0.290	2.580	0.275
T_{2a}^*	−2.320	0.020	−1.752	0.080	0.720	0.483
T_{2b}^*	1.721	0.085	3.799	0.001	−1.712	0.087
T_{2c}^*	−2.767	0.006	−2.452	0.004	0.935	0.350
Goodness-of-fit testing of linear submodel, $H_0^* : \beta_i(t) = \eta_1 + \eta_2 t$						
T_1^*	4.456	0.108	0.688	0.709		
T_{2a}^*	−0.684	0.494	−0.517	0.605		
T_{2b}^*	0.399	0.690	0.053	0.958		
T_{2c}^*	−1.265	0.206	−0.726	0.468		

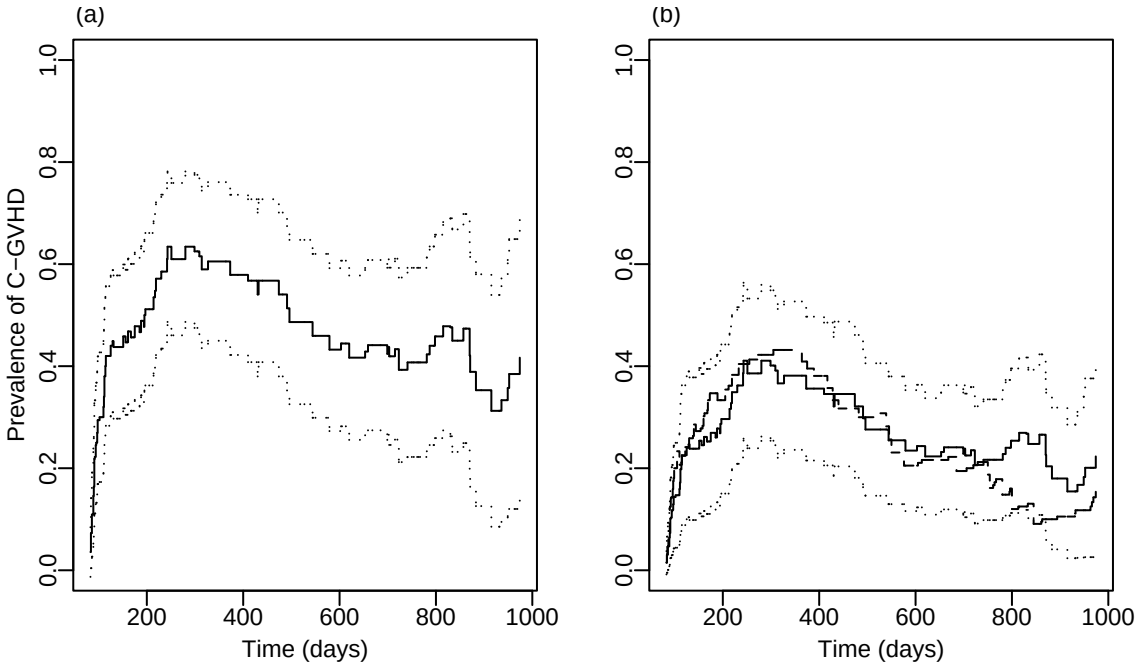

Figure 1. Estimates of prevalence of C-GVHD. The solid line is the nonparametric estimate, the dashed line is the parametric estimate, and the dotted line is the 0.95 pointwise confidence intervals from the parametric estimate. (a) MC, (b) PMC.

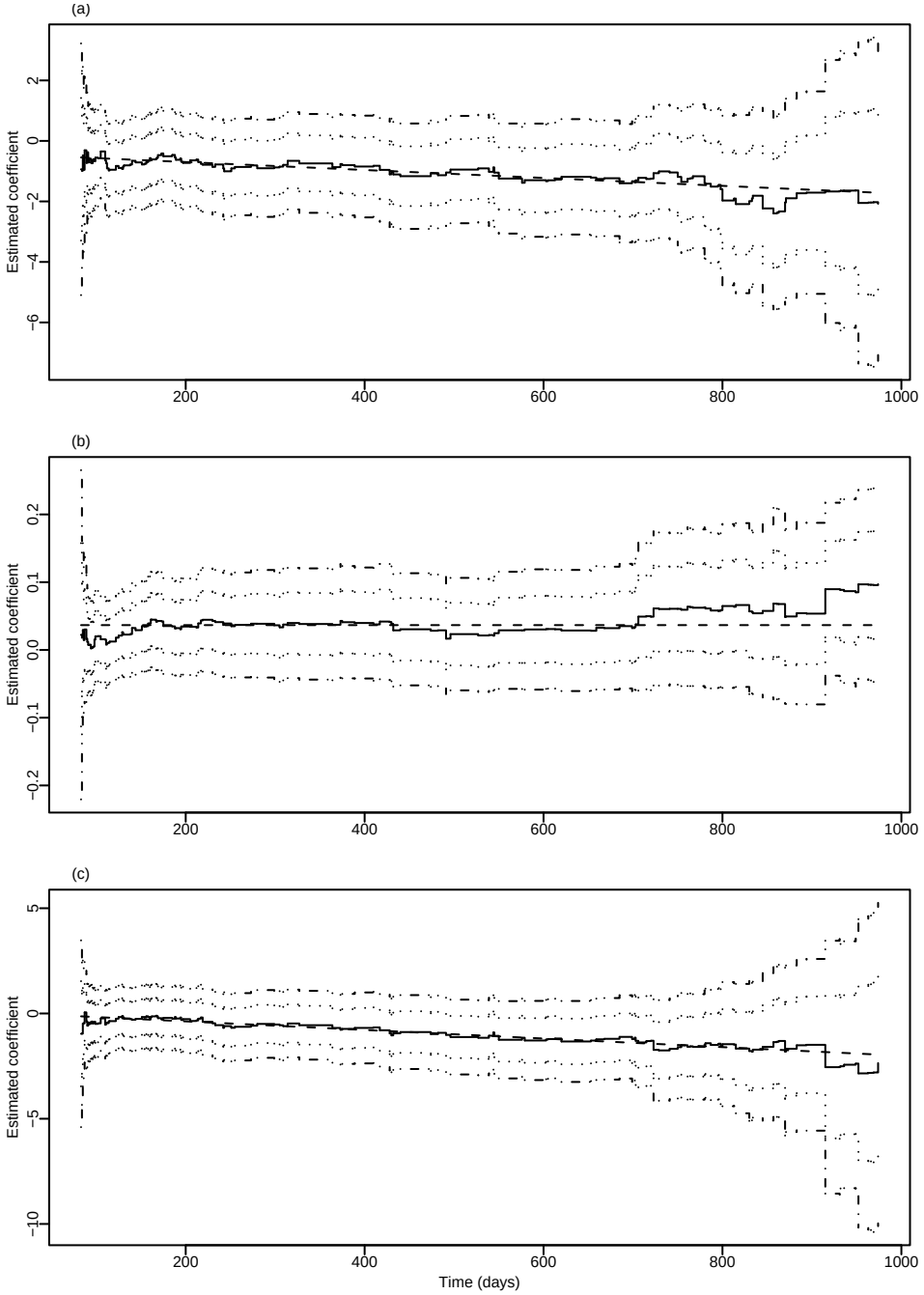

Figure 2. Estimates of regression coefficients. The solid line is the nonparametric estimate, the dashed line is the parametric estimate, the two innermost and two outermost dotted lines are 0.95 pointwise confidence intervals and 0.95 simultaneous confidence bands from the nonparametric estimate, respectively. (a) treatment, (b) age, (c) gender.