1-27-2003

The Sensitivity and Specificity of Markers for Event Times

Tianxi Cai
Harvard University, tcai@hsph.harvard.edu

Margaret S. Pepe
University of Washington, mspepe@u.washington.edu

Thomas Lumley
University of Washington, tlumley@u.washington.edu

Yingye Zheng
University of Washington, zheng@u.washington.edu

Nancy Swords Jenny
University of Vermont, Nancy.Jenny@uvm.edu

1. Introduction

The use of clinical and laboratory data to predict future patient events is a very popular idea in medicine at present. Biomarkers are under development to detect cancer before onset of clinical disease (Pepe et al, 2001). Gene expression profiles of tumor tissue promise to be predictive of survival in cancer patients (Veer et al, 2002). Clinical scores, such as the Framingham Risk Score (Wilson et al, 1998), are considered predictive of myocardial infarction and stroke. It is critical to evaluate the sensitivity and specificity of such predictors or markers before adopting them for use in clinical practice.

The literature on evaluating the accuracy of a marker, predictor or diagnostic test result, Y, deals primarily with settings where Y is measured concurrently with the gold standard disease variable D. The True Positive Rate (TPR) and False Positive Rate (FPR) functions are:

$$\text{TPR}(y) = \text{pr}(Y \geq y \mid D = 1), \qquad \text{FPR}(y) = \text{pr}(Y \geq y \mid D = 0),$$

where $D = 1$ indicates disease present, $D = 0$ denotes its absence and the threshold y is used to define a positive test result as $Y \geq y$. If Y is continuous, an ROC curve that plots $\text{TPR}(y)$ versus $\text{FPR}(y)$ for all possible values of y is often used to describe the discriminatory capacity of Y.

The notions of true and false positive rates must be extended when the outcome is an event time random variable T and the time at which Y is measured relative to T can vary. Indeed, the timing of the measurement, denoted by s, is likely to impact on the capacity of $Y(s)$ to predict T. Measurements made closer to the event time are likely to be more predictive than those made earlier. We define the time dependent TPR and FPR functions:

$$\text{TPR}_{t,s}(y) = \text{pr}\left\{Y(s) \geq y \mid T = t\right\}, \qquad \text{FPR}_{\tau,s}(y) = \text{pr}\left\{Y(s) \geq y \mid T > \tau\right\}, \tag{1}$$

where $0 \leq t \leq \tau$ and τ is some suitably chosen large time point. The FPR definition implicitly assumes that subjects without events by τ are an adequate control group against the cases which occur over the time interval $[0, \tau]$ are compared. These definitions are consistent with those used by Balasubramanian and Lagakos (2001), Leisenring, Pepe and Longton (1997) and Etzioni et al (1999). Heagerty, Pepe and Lumley (2000) define cumulative incidence based TPR and FPR functions:

$$\text{TPR}_{t,s}^{\text{CI}}(y) = \text{pr}\left\{Y(s) \geq y \mid T \leq t\right\}, \qquad \text{FPR}_{t,s}^{\text{CI}}(y) = \text{pr}\left\{Y(s) \geq y \mid T > t\right\}. \tag{2}$$

2

However, the definitions in (1) lead to more straightforward regression modeling procedures and are easier to interpret as functions of time than their cumulative incidence based counterparts (Pepe, 2003). Moreover, we can calculate (2) from (1) with knowledge of the event time distribution. Thus we concentrate on (1) here initially, and return to (2) in the example.

In this paper we consider models for the time dependent (TPR, FPR) functions and procedures to make inference about them from prospective cohort studies. The marker Y may be measured at multiple times for a subject, covariates $\mathbf{Z}$ that affect the true and false positive rates may be available and the event time T can be right censored. We extend the marginal regression modeling approach of Leisenring, Pepe and Longton (1997) which deals only with binary markers Y and uncensored failure times. The models and assumptions are described in the next section. Inference about unknown parameters and about the time dependent (TPR, FPR) functions is discussed in sections 3 and 4. Simulation studies described in section 5 suggest that the procedures work well in finite samples when the assumptions hold. In section 6 we apply the methods to data from the Cardiovascular Health Study, a prospective cohort study of older adults (Fried et al, 1991). We investigate the sensitivity (TPR) and specificity (1-FPR) of the Framingham risk score as a marker for future cardiovascular events in this population. As expected, we show that the score is better at discriminating short term than long term risk and works better in females than in males. However, the score is not a very accurate predictor in any subgroup studied. We close in section 7 with a discussion of alternative approaches to the evaluation of markers for event time data.

2.   Models and Assumptions

Since the purpose is prediction, we only consider marker measurements made prior to the event time, $Y(s)$ for $s \leq T$. The event time can be right censored and we assume random censorship, where the observation time $X = \min(T, C)$ for a censoring variable C that is independent of T and Y conditional on the covariates. The censoring indicator $\delta = 1$ if T is uncensored, i.e. $T \leq C$, while $\delta = 0$ otherwise.

Suppose that the data for analysis are organized as N data records for n subjects:

$$\{(Y_{ik}, s_{ik}, \mathbf{Z}_i, X_i, \delta_i), \ k = 1, \dots, K_i, \ i = 1, \dots, n\},$$

where Y_{ik} is the biomarker value for the ith subject at the kth measurement time s_{ik} and $\mathbf{Z}_i$ is the

3

$p \times 1$ covariate vector. Consider the following marginal probability models:

$$\text{TPR}_{t,s_{ik},\mathbf{Z}_i}(y) = \text{pr}(Y_{ik} \geq y \mid T_i = t, \mathbf{Z}_i, s_{ik}) = g\left\{\boldsymbol{\eta}_{\boldsymbol{\alpha}_0}(t, s_{ik}) + \boldsymbol{\beta}_0'\mathbf{Z}_i + h_0(y)\right\}, \quad t \in [s_{ik}, \tau], \quad (3)$$

$$\text{FPR}_{\tau,s_{ik},\mathbf{Z}_i}(y) = \text{pr}(Y_{ik} \geq y \mid T_i > \tau, \mathbf{Z}_i, s_{ik}) = g_\tau\left\{\xi_{\boldsymbol{a}_0}(s_{ik}) + \mathbf{b}_0'\mathbf{Z}_i + c_0(y)\right\}, \quad (4)$$

where g, g_τ are specified inverse link functions, h_0 and c_0 are baseline functions of the threshold y that are completely unspecified, and the dependence on time is through the parametric functions $\eta_{\boldsymbol{\alpha}_0}(t, s) = \boldsymbol{\alpha}_0'\boldsymbol{\eta}(t, s)$, $\xi_{\boldsymbol{a}_0}(s) = \boldsymbol{a}_0'\boldsymbol{\xi}(s)$, where $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are vectors of polynomial or spline basis functions. One possibility for $\boldsymbol{\eta}(t, s)$ is to specify it as a function of $t - s$, which is to say that the sensitivity $\text{TPR}_{t,s}(\cdot)$ of the marker depends only on the time lag between the marker measurement and the event. However, in some applications the absolute time of the marker measurement from the origin may also affect the TPR functions. Examples include settings where $s = 0$ denotes entry into an intervention study or if s denotes the subject's age which is associated with the marker distribution. In the special case where s_{ik} is a constant for all observations (e.g. $s_{ik} = 0$), we set $\xi_{\boldsymbol{a}_0}(s) = 0$.

The non-parametric baseline functions of y, $h_0(\cdot)$ and $c_0(\cdot)$, essentially define the shape and location of the sensitivity and specificity functions, while the parameters $\boldsymbol{\beta}_0$ and $\mathbf{b}_0$ quantify covariate effects on them. When Y is binary, we see that $h_0(\cdot)$ and $c_0(\cdot)$ are constants. The marginal logistic regression models proposed by Leisenring, Pepe and Longton (1997) for binary Y correspond to (3) and (4) with logit link functions. Note that, when Y is continuous, this type of model corresponds to the marginal semi-parametric transformation model (Dabrowska & Doksum 1988a, 1988b; Cheng, Wei & Ying 1995, 1997; Scharfstein, Tsiatis & Gilbert 1998; Cai, Wei & Wilcox, 2000). That is, models (3) and (4) can be represented as:

$$h_0(Y_{ik}) = -\boldsymbol{\eta}_{\boldsymbol{\alpha}_0}(T_i, s_{ik}) - \boldsymbol{\beta}_0'\mathbf{Z}_i + \epsilon_{ik}, \quad \text{if } T_i \in [s_{ik}, \tau],$$

$$c_0(Y_{ik}) = -\xi_{\boldsymbol{a}_0}(s_{ik}) - \mathbf{b}_0'\mathbf{Z}_i + \varepsilon_{ik}, \quad \text{if } T_i > \tau,$$

where $\text{pr}(\epsilon_{ik} \geq y) = g(y)$ and $\text{pr}(\varepsilon_{ik} \geq y) = g_\tau(y)$.

In order to incorporate data from subjects who are censored before τ, we need to model the event time distribution. Suppose that a proportional hazards model holds:

$$\lambda(t \mid \mathbf{Z}_i) = \lambda_0(t)\exp(\boldsymbol{\gamma}_0'\mathbf{Z}_i), \quad (5)$$

4

where $\lambda(t \mid \mathbf{Z}_i)$ is the hazard function for the ith subject and the baseline hazard function $\lambda_0(t)$ is unspecified. It follows that

$$S_{\mathbf{Z}_i}(t) = \mathrm{pr}(T \geq t \mid \mathbf{Z}_i) = \exp\left\{-\Lambda_0(t)\exp(\boldsymbol{\gamma}_0'\mathbf{Z}_i)\right\},$$

where $\Lambda_0(\cdot)$ is the baseline cumulative hazard function. Other regression models for T can be used. Our estimation procedure only requires a consistent estimate of $S_{\mathbf{z}}(t)$ and does not rely on the proportional hazards model assumption.

3. Estimating the Model Components

In this section, we derive estimation procedures for $\boldsymbol{\psi}_0 = \left\{\mathbf{H}_0(\cdot) = [h_0(\cdot), c_0(\cdot)]', \boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{10}', \boldsymbol{\theta}_{20}']'\right\}$, where $\boldsymbol{\theta}_{10} = [\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0']'$, $\boldsymbol{\theta}_{20} = [\mathbf{a}_0', \mathbf{b}_0']'$. To this end, let $\Delta_i = I(X_i \leq \tau)(1 - \delta_i)$ denote whether the ith subject is censored before τ and $\mathcal{F}_{ik} = \{I(X_i \leq \tau), \delta_i, \min(X_i, \tau), s_{ik}, \mathbf{Z}_i\}$. Without loss of generality, we index observations such that $X_i \leq \tau$ and $\delta_i = 1$ with $i = 1, \ldots, n_d$; $X_i \leq \tau$ and $\delta_i = 0$ for $i = n_d + 1, \ldots, n_\tau$; $X_i > \tau$ for $i = n_\tau + 1, \ldots, n$. We can think of the first and last groups as the cases and controls, respectively, while the case/control status of the second group, i.e. those censored in $(0, \tau)$, are unknown.

We base inference on the indicator variable $I(Y_{ik} \geq y)$ since models (3) and (4) essentially relate the conditional expectation $p_{ik}(y; \boldsymbol{\psi}_0) = E\left\{I(Y_{ik} \geq y) \mid \mathcal{F}_{ik}; \boldsymbol{\psi}_0\right\}$ to the parameters of interest $\boldsymbol{\psi}_0$. To estimate $\mathbf{H}_0(y)$, we mimic the score equation based on a binomial likelihood and consider the following estimating equation:

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \mathbf{W}_{ik}^{\mathbf{H}}(y; \boldsymbol{\psi})\left\{I(Y_{ik} \geq y) - p_{ik}(y; \boldsymbol{\psi})\right\} = 0, \qquad \text{for each } y \in [\mathfrak{l}, \mathfrak{u}], \tag{6}$$

where

$$\mathbf{W}_{ik}^{\mathbf{H}}(y; \boldsymbol{\psi}) = \frac{\rho^{\Delta_i} \frac{\partial}{\partial \mathbf{H}} p_{ik}(y; \boldsymbol{\psi})}{p_{ik}(y; \boldsymbol{\psi})\{1 - p_{ik}(y; \boldsymbol{\psi})\}},$$

ρ is a pre-specified non-negative weight, and $\mathfrak{l}, \mathfrak{u}$ are pre-determined constants such that $\mathrm{pr}(Y_{ik} < \mathfrak{l})$ and $\mathrm{pr}(Y_{ik} > \mathfrak{u})$ are both positive. To estimate $\boldsymbol{\theta}_0$, we propose to solve

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{\mathfrak{l}}^{\mathfrak{u}} \mathbf{W}_{ik}^{\boldsymbol{\theta}}(y; \boldsymbol{\psi})\left\{I(Y_{ik} \geq y) - p_{ik}(y; \boldsymbol{\psi})\right\}d\hat{v}(y) = 0, \tag{7}$$

5

where

$$\mathbf{W}_{ik}^{\boldsymbol{\theta}}(y;\boldsymbol{\psi}) = \frac{\rho^{\Delta_i} \frac{\partial}{\partial \boldsymbol{\theta}} p_{ik}(y;\boldsymbol{\psi})}{p_{ik}(y;\boldsymbol{\psi})\{1 - p_{ik}(y;\boldsymbol{\psi})\}},$$

$\hat{v}(\cdot)$ is some increasing function that can depend on the data but converges asymptotically to a deterministic function $v(y)$ uniformly in $y \in [\mathfrak{l}, \mathfrak{u}]$. The basic idea then is to solve (6) and (7) simultaneously to estimate the parameters in the models (3) and (4).

The binomial probability $p_{ik}(y;\boldsymbol{\psi})$ depends on the case/control/censored status of the observation. To derive the form of $p_{ik}(y;\boldsymbol{\psi})$, we first note that $\mathrm{pr}(Y_{ik} \geq y \mid \mathcal{F}_{ik}) = \mathrm{pr}(Y_{ik} \geq y \mid T_i, \mathbf{Z}_i, s_{ik})$ for $1 \leq i \leq n_d$ and $\mathrm{pr}(Y_{ik} \geq y \mid \mathcal{F}_{ik}) = \mathrm{pr}(Y_{ik} \geq y \mid T_i > \tau, \mathbf{Z}_i, s_{ik})$ for $n_\tau < i \leq n$. If the ith subject is censored at $X_i = x < \tau$, we write

$$\begin{aligned}
\mathrm{pr}(Y_{ik} \geq y \mid \mathcal{F}_{ik}) &= \frac{\mathrm{pr}(Y_{ik} \geq y, \tau \geq T_i > x \mid \mathbf{Z}_i, s_{ik}) + \mathrm{pr}(Y_{ik} \geq y, T_i > \tau \mid \mathbf{Z}_i, s_{ik})}{S_{\mathbf{Z}_i}(x)} \\
&= \frac{-\int_x^\tau \mathrm{pr}(Y_{ik} \geq y \mid T_i = t, \mathbf{Z}_i, s_{ik}) dS_{\mathbf{Z}_i}(t) + \mathrm{pr}(Y_{ik} \geq y, T_i > \tau \mid \mathbf{Z}_i, s_{ik})}{S_{\mathbf{Z}_i}(x)}.
\end{aligned}$$

which involves the unknown function $S_{\mathbf{Z}_i}(\cdot)$. It follows from the models (3) and (4) that

$$p_{ik}(y;\boldsymbol{\psi}) = \begin{cases} g\left\{\boldsymbol{\theta}_1' \vec{\mathbf{Z}}_{ik} + h(y)\right\}, & 1 \leq i \leq n_d \\ \dfrac{-\int_{X_i}^\tau g\left\{\boldsymbol{\theta}_1' \vec{\mathbf{Z}}_{ik}^{[u]} + h(y)\right\} dS_{\mathbf{Z}_i}(u) + g_\tau\{\boldsymbol{\theta}_2' \mathbf{Z}_{ik} + c(y)\} S_{\mathbf{Z}_i}(\tau)}{S_{\mathbf{Z}_i}(X_i)}, & n_d < i \leq n_\tau \\ g_\tau\left\{\boldsymbol{\theta}_2' \mathbf{Z}_{ik} + c(y)\right\}, & n_\tau < i \leq n \end{cases}$$

where

$$\vec{\mathbf{Z}}_{ik} = \begin{bmatrix} \boldsymbol{\eta}(X_i, s_{ik}) \\ \mathbf{Z}_i \end{bmatrix}, \quad \vec{\mathbf{Z}}_{ik}^{[u]} = \begin{bmatrix} \boldsymbol{\eta}(u, s_{ik}) \\ \mathbf{Z}_i \end{bmatrix}, \quad \mathbf{Z}_{ik} = \begin{bmatrix} \boldsymbol{\xi}(s_{ik}) \\ \mathbf{Z}_i \end{bmatrix}.$$

Observe that the estimating equations include a weighting factor ρ that dictates the extent to which the censored observations in $(0, \tau)$ enters into the analysis (we have $\rho^{\Delta_i} = 1$ for known cases and controls because $\Delta_i = 0$ for them). If ρ is set to 0, then censored observations are excluded entirely from the analysis and the equations correspond to Leisenring et al (1997). Increasing ρ allows censored observations to have more influence on estimation. A variety of values for ρ are investigated in the simulation studies. For now note that, when $\rho > 0$, an estimate of the survivor function $S_{\mathbf{Z}_i}(\cdot)$ is required in order to approximate the probabilities $p_{ik}(y;\boldsymbol{\psi})$ for censored observations. As well, to obtain an estimate of the cumulative incidence based TPR$^{\mathrm{CI}}$ function, one needs to use an estimate of $S_{\mathbf{z}}(t)$.

6

We propose a three step approach to estimating $\boldsymbol{\psi}_0$: 1) estimation of $S_{\mathbf{z}}(t)$; 2) estimation of $\boldsymbol{\psi}_0$ when $\rho = 0$ and 3) estimation of $\boldsymbol{\psi}_0$ when $\rho > 0$. We detail the three steps next.

3.1 Estimation of $S_{\mathbf{z}}(t)$

Under the Cox proportional hazards model (5), we estimate $S_{\mathbf{Z}}(t)$ as

$$\widehat{S}_{\mathbf{Z}}(t) = \exp\{-\hat{\Lambda}_0(t)\exp(\hat{\boldsymbol{\gamma}}'\mathbf{Z})\},$$

where $\hat{\boldsymbol{\gamma}}$ is obtained by maximizing the partial likelihood based on data $\{X_i, \delta_i, \mathbf{Z}_i; \ i = 1, \dots, n\}$, and $\hat{\Lambda}_0(t)$ is the Breslow estimate of the cumulative baseline hazard function $\Lambda_0(t)$ (Fleming and Harrington, 1991).

3.2 Estimation of $\boldsymbol{\psi}_0$ when $\rho = 0$

Let $\widetilde{\boldsymbol{\psi}} = \left\{\widetilde{\mathbf{H}}(\cdot), \widetilde{\boldsymbol{\theta}}\right\}$ denote the solution to the estimating equations (6) and (7) when $\rho = 0$, i.e. when the censored observations $i = n_d + 1, \dots, n_\tau$ are ignored. We show in Appendix A that $\widetilde{\boldsymbol{\psi}}$ is a consistent estimator of $\boldsymbol{\psi}_0$ for $y \in [\mathfrak{l}, \mathfrak{u}]$. An estimate of $S_{\mathbf{z}}(\cdot)$ is not required to obtain the estimator $\widetilde{\boldsymbol{\psi}}$. $\{\boldsymbol{\theta}_{10}, h_0(\cdot)\}$ are estimated using only the cases, subjects $i = 1, \dots, n_d$, and $\{\boldsymbol{\theta}_{20}, c_0(\cdot)\}$ are estimated using only the controls, i.e. subjects $i = n_\tau + 1, \dots, n$.

3.3 Estimation of $\boldsymbol{\psi}_0$ when $\rho \neq 0$

To include subjects $i = n_d + 1, \dots, n_\tau$ in estimating $\boldsymbol{\psi}_0$, we set $\rho \neq 0$ and approximate $p_{ik}(y; \boldsymbol{\psi})$ for those censored subjects by estimating the unknown survivor function $S_{\mathbf{z}}(\cdot)$. Let $\widehat{p}$, $\widehat{\mathbf{W}}_{ik}^{\mathbf{H}}$, $\widehat{\mathbf{W}}_{ik}^{\boldsymbol{\theta}}$ be p, $\mathbf{W}_{ik}^{\mathbf{H}}$, $\mathbf{W}_{ik}^{\boldsymbol{\theta}}$, with $S_{\mathbf{Z}_i}(\cdot)$ replaced with $\widehat{S}_{\mathbf{Z}_i}(\cdot)$, respectively. In analogy with (6), we consider the following estimating equation for $\mathbf{H}(y)$ at a given $\boldsymbol{\theta}$:

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \widehat{\mathbf{W}}_{ik}^{\mathbf{H}}(y; \widetilde{\boldsymbol{\psi}})\left\{I(Y_{ik} \geq y) - \widehat{p}_{ik}(y; \boldsymbol{\psi})\right\} = 0, \qquad \text{for any } y \in [\mathfrak{l}, \mathfrak{u}]. \tag{8}$$

Let $\widehat{\mathbf{H}}(y; \boldsymbol{\theta}) = [\widehat{h}(y; \boldsymbol{\theta}), \widehat{c}(y; \boldsymbol{\theta})]'$ denote the solution to (8) given $\boldsymbol{\theta}$ and let $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}} = \left\{\widehat{\mathbf{H}}(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta}\right\}$. To estimate $\boldsymbol{\theta}_0$, we use

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{\mathfrak{l}}^{\mathfrak{u}} \widehat{\mathbf{W}}_{ik}^{\boldsymbol{\theta}}(y; \widetilde{\boldsymbol{\psi}})\left\{I(Y_{ik} \geq y) - \widehat{p}_{ik}(y; \widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}})\right\} d\hat{v}(y) = 0. \tag{9}$$

7

Observe that the weight functions $\widehat{\mathbf{W}}_{ik}^{\mathbf{H}}(y; \boldsymbol{\psi})$ and $\widehat{\mathbf{W}}_{ik}^{\boldsymbol{\theta}}(y; \boldsymbol{\psi})$ are evaluated at $\boldsymbol{\psi} = \widetilde{\boldsymbol{\psi}}$ and are not updated with $\boldsymbol{\psi}_{\boldsymbol{\theta}}$. This results in a much simpler Jacobian matrix and reduces the computational requirements substantially.

In summary, the strategy for estimating $\boldsymbol{\psi}_0$ is to solve (6) and (7) with ρ set to 0 in order to derive an initial estimate $\widetilde{\boldsymbol{\psi}}$. This, together with an estimate of $S_{\mathbf{Z}_i}(\cdot)$ is substituted into equations (6) and (7) with ρ set to some positive value to incorporate the censored observations in $[0, \tau)$. Then final estimates of $\boldsymbol{\psi}_0$ are calculated by solving (8) and (9).

4. Inference in Large Samples

4.1 *Asymptotic properties of* $\widehat{\boldsymbol{\theta}}$

Let $\widehat{\boldsymbol{\theta}}$ be the solution to (9), $\widehat{h}(y) = \widehat{h}(y; \widehat{\boldsymbol{\theta}})$ and $\widehat{c}(y) = \widehat{c}(y; \widehat{\boldsymbol{\theta}})$. The following results are shown in appendices A and B

Theorem 1. $\left\{ \widehat{\boldsymbol{\theta}}, \widehat{h}(y), \widehat{c}(y) \right\}$ *are unique for large n and are consistent.*

Theorem 2. $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *is asymptotically equivalent to*

$$n^{-\frac{3}{2}} \mathbb{A}^{-1} \sum_{i=n_d+1}^{n_\tau} \sum_{j=1}^{n} \mathbf{U}_{ij} \tag{10}$$

where

$$\mathbf{U}_{ij} = \int_{\{}^{u} \left\{ \mathcal{E}_{ij}(y; \mathbf{v}^{\boldsymbol{\theta}}) + \sum_{l=1}^{K_j} \mathbf{v}_{jl}^{\boldsymbol{\theta}}(y) e_{jl}(y) \right\} dv(y),$$

$$\mathcal{E}_{ij}(y; f) = \sum_{k=1}^{K_i} f_{ik}(y) \left[\int_{X_i}^{\tau} g \left\{ \boldsymbol{\theta}_{10}' \vec{\mathbf{Z}}_{ik}^{[u]} + h_0(y) \right\} d\mathfrak{s}_j(u, X_i; \mathbf{Z}_i) + g_\tau \left\{ \boldsymbol{\theta}_{20}' \mathbf{Z}_{ik} + c_0(y) \right\} \mathfrak{s}_j(\tau, X_i; \mathbf{Z}_i) \right],$$

$\mathbb{A}$ *is defined in (A.6),* $e_{ik}(y) = I(Y_{ik} \geq y) - p_{ik}(y; \boldsymbol{\psi}_0)$, $\mathbf{v}_{jl}^{\boldsymbol{\theta}}(y)$ *is defined in (B.7) and* $\mathfrak{s}_j(u, x; \mathbf{z})$ *are defined in (B.8).*

The latter result is used to obtain interval estimates of specific components of $\boldsymbol{\theta}_0$. In particular, note that (10) is a U-statistics (Serfling, 1980) with covariance matrix

$$\boldsymbol{\Sigma} = n^{-3} \left\{ \sum_{\substack{n_d < i \neq i' \leq n_\tau \\ 1 \leq j \leq n}} \mathbf{U}_{ij} \mathbf{U}_{i'j}' + \sum_{\substack{n_d < i \leq n_\tau \\ 1 \leq j \neq j' \leq n}} \mathbf{U}_{ij} \mathbf{U}_{ij'}' + 2 \sum_{\substack{n_d < i \neq i' \leq n_\tau \\ 1 \leq j \leq n}} \mathbf{U}_{ij} \mathbf{U}_{i'i}' \right\}$$

8

Now, let $\widehat{\mathbb{A}}$ and $\widehat{\Sigma}$ be the matrices obtained by replacing all the theoretical quantities in $\mathbb{A}$ and Σ with their empirical counterparts. Then, the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be approximated by a normal distribution with mean 0 and covariance matrix $\widehat{\mathbb{A}}^{-1} \widehat{\Sigma} \widehat{\mathbb{A}}^{-1}$.

4.2 *Estimating the time dependent sensitivity and specificity functions*

We now turn to inference about the TPR and FPR functions, which depend on time and on co-variates $\mathbf{z}$. Substituting $\left\{ \widehat{\boldsymbol{\theta}}, \widehat{h}(y), \widehat{c}(y) \right\}$ into the model forms (3) and (4) we propose estimates for $\text{TPR}_{t,s,\mathbf{z}}(y)$ and $\text{FPR}_{\tau,s,\mathbf{z}}(y)$ as:

$$\widehat{\text{TPR}}_{t,s,\mathbf{z}}(y) = g \left\{ \eta_{\widehat{\boldsymbol{\alpha}}}(t,s) + \widehat{\boldsymbol{\beta}}' \mathbf{z} + \widehat{h}(y) \right\} ,$$

$$\widehat{\text{FPR}}_{\tau,s,\mathbf{z}}(y) = g_\tau \left\{ \xi_{\widehat{\boldsymbol{a}}}(s) + \widehat{\mathbf{b}}' \mathbf{z} + \widehat{c}(y) \right\} .$$

Asymptotic distribution theory for these functions is presented in appendix B where we prove the following theorem:

Theorem 3. $\mathbf{Q}(y; t, s, \mathbf{z}) = n^{\frac{1}{2}} \begin{bmatrix} g^{-1} \left\{ \widehat{\mathit{TPR}}_{t,s,\mathbf{z}}(y) \right\} - g^{-1} \left\{ \mathit{TPR}_{t,s,\mathbf{z}}(y) \right\} \\ g_\tau^{-1} \left\{ \widehat{\mathit{FPR}}_{\tau,s,\mathbf{z}}(y) \right\} - g_\tau^{-1} \left\{ \mathit{FPR}_{\tau,s,\mathbf{z}}(y) \right\} \end{bmatrix}$ *is asymptotically equivalent to*

$$n^{-\frac{3}{2}} \sum_{i=n_d+1}^{n_\tau} \sum_{j=1}^{n} \left\{ \mathcal{H}_{ij}(y) + \begin{bmatrix} \boldsymbol{\eta}(t,s)' & \mathbf{z}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\xi}(s)' & \mathbf{z}' \end{bmatrix} \mathbb{A}^{-1} \mathbf{U}_{ij} \right\} ,$$

where $\mathcal{H}_{ij}(y)$ is defined in (B.9).

Theorem 3 allows us to approximate the distribution of the process $\mathbf{Q}(y; t, s, \mathbf{z})$ using re-sampling techniques (Parzen, Wei and Ying, 1994) in practice. A detailed description of the procedure for constructing confidence bands based on the re-sampling method can be found in Cai & Pepe (2002).

We mentioned earlier that alternative definitions for time dependent TPR and FPR are possible. Estimators for the cumulative incidence based TPR and FPR functions can be derived from ours

9

noting the following identities:

$$\text{TPR}^{\text{CI}}_{t,s,\mathbf{z}}(y) = \text{pr}(Y(s) \geq y \mid T \leq t, \mathbf{z})$$

$$= \frac{-\int_s^t g\left\{\eta\boldsymbol{\alpha}_0(u,s) + \boldsymbol{\beta}_0'\mathbf{z} + h_0(y)\right\} dS_\mathbf{z}(u)}{1 - S_\mathbf{z}(t)},$$

$$\text{FPR}^{\text{CI}}_{t,s,\mathbf{z}}(y) = \text{pr}(Y(s) \geq y \mid T > t, \mathbf{z})$$

$$= \frac{-\int_s^t g\left\{\eta\boldsymbol{\alpha}_0(u,s) + \boldsymbol{\beta}_0'\mathbf{z} + h_0(y)\right\} dS_\mathbf{z}(u) + g_\tau\left\{\xi\boldsymbol{a}_0(s) + \mathbf{b}_0'\mathbf{z} + c(y)\right\} S_\mathbf{z}(\tau)}{S_\mathbf{z}(t)}.$$

Plugging in $\left\{\widehat{\boldsymbol{\theta}}, \widehat{h}(\cdot), \widehat{c}(\cdot)\right\}$ for $\boldsymbol{\psi}_0$, we obtain consistent estimators for $\text{TPR}^{\text{CI}}_{t,s,\mathbf{z}}(y)$ and $\text{FPR}^{\text{CI}}_{t,s,\mathbf{z}}(y)$.

5. Simulation Studies

We performed a simulation study to illustrate the estimation procedure proposed in the previous sections and to investigate the impact of ρ on their efficiencies. We simulated data as follows: a covariate Z with Uniform$(0, 1)$ distribution was generated for each of $n = 200$ individuals. Survival times were generated from a proportional hazards regression model with regression co-efficient $\gamma_0 = -0.5$ and baseline hazard function being a mixture of two Weibull distributions, Weibull$(1, 2)$ with probability 0.3 and Weibull$(2, 11)$ with probability 0.7. We chose $\tau = 5$ so that $\text{pr}(T > \tau) = 0.72$. The censoring random variable C also had a mixture distribution being de-rived from a Weibull$(2, \tau/4)$ with probability $p_c = 0.3$ and otherwise $C = 10$. Thus about 28% of subjects are censored before τ of which most, 22% have $T > \tau$, and 50% of subjects remain under observation and event free at τ.

Two marker measurements were generated for each individual, one at $s_{i1} = 0$ and one at a random time s_{i2} uniformly distributed in the time interval $[0.5, 1.5]$. We discarded observations Y_{ik} if they were measured at times beyond their event or censoring time. The marker measurements themselves were generated for the cases (i.e. $T_i \leq \tau$) from the model

$$h(Y_{ik}) = -\frac{1}{2}(T_i - s_{ik}) - Z_i + \epsilon_{ik},$$

with $h(y) = 3\log(y - 2) - 2$ and $(\epsilon_{i1}, \epsilon_{i2})$ having a standard bivariate normal distribution with covariance 0.2. For the controls $(T_i > \tau)$ the model was

$$c(Y_{ik}) = -Z_i + \varepsilon_{ik},$$

10

with $c(y) = 2\log(y-2) + 1$ and $(\varepsilon_{i1}, \varepsilon_{i2})$ having a standard bivariate normal distribution with covariance 0.2. These models correspond to the following time-dependent TPR and FPR functions:

$$\text{TPR}_{t,s,z}(y) = 1 - \Phi\left\{\frac{1}{2}(t-s) + z + 3\log(y-2) - 2\right\},$$

$$\text{FPR}_{\tau,s,z}(y) = 1 - \Phi\left\{z + 2\log(y-2) + 1\right\},$$

where Φ denotes the cumulative normal distribution function. These are shown in Figure 1. Thus the parameters $(\alpha_0, \beta_0) = (\frac{1}{2}, 1)$ and $b_0 = 1$ in this setting while the functions $\eta(t,s) = t-s$ and $\xi(s) = 0$. The inverse link functions are $g(\cdot) = g_\tau(\cdot) = 1 - \Phi(\cdot)$.

In Table 1 (a), we present the empirical bias and mean squared error of $\widehat{\boldsymbol{\theta}}$ estimated using values of $\rho = 0, 0.5, 1.0, 2.0$ and 5.0. In Figures 1 and 2, we present the sample average of the estimated TPR and FPR for $Y(0)$ at event time $t = 2$ as functions of y and as functions of event time t using the fixed threshold $y = 3$. The results shown in Table 1(a) and the figures suggest that our method provides reasonably unbiased estimates of both the model parameters and (TPR, FPR) functions. The estimators have smaller standard errors when markers are measured at two time points, $K = 2$, than when only the baseline measurements at $s = 0$ are available, $K = 1$. This makes sense since we expect to gain efficiency by using more information.

The choice of ρ affects the efficiency of all parameters of interest, but seems to have most impact on $\widehat{b}$, less on $\widehat{\beta}$ and very little on $\widehat{\alpha}$ in this setting. Recall that when $\rho = 0$, censored observations are ignored for estimation. By using $\rho > 0$, we find that the estimates are almost always more precise than those calculated with $\rho = 0$. In particular, the efficiencies for estimating (α, β, b) of $\widetilde{\boldsymbol{\theta}}$ which is calculated with $\rho = 0$ relative to $\widehat{\boldsymbol{\theta}}$ calculated using $\rho = 1$ are (0.98, 0.96, 0.79), respectively, when $K = 1$. The results vary across settings, however, with efficiency gained for one parameter often accompanied by a loss for another parameter. We also found that the amount of efficiency gain using $\rho > 0$ varies with the censoring pattern, but no general pattern appeared to emerge.

The small gain in efficiency for $\widehat{\alpha}$ and $\widehat{\beta}$ by using $\rho > 0$ can in part be attributed to the low percentage of censored subjects with $T \leq \tau$. In another simulation, we generated T from the Cox model with $\Lambda_0(t)$ being the cumulative hazard function for the Weibull mixture of Weibull$(2, 4)$ with probability 0.5 and Weibull$(5, 15)$ with probability 0.5, and let $\mathcal{T} = 15$, $\tau = 5$, $p_c = 0.5$. This configuration results in about 50% of subjects censored before τ (of which 12% have $T \leq \tau$ and 38% have $T > \tau$), and 37% of subjects under observation and event free at τ. There is more to be

11

gained in this setting from including censored data in the analysis (Table 1(b)). For example, the relative efficiencies for estimating (α, β, b) with $\rho = 0$ compared with $\rho = 1$ are (0.98, 0.88, 0.72).

How should the weight of censored data be chosen in practice? To select an optimal ρ, one could use an "optimization procedure" to minimize the total mean squared error $E\{\mathbf{1}'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^2\}$. That is, one selects the value of ρ that corresponds to the smallest value for the estimated sum of square errors. Using this criterion, in our simulations the weight of choice is $\rho = 1$ in both configurations. Alternatively, one could minimize other quantities such as the total coefficient of variation $\sum_{j=1}^{P} \sqrt{\mathrm{var}(\widehat{\theta}_j)/\widehat{\theta}_j^2}$, where P is the dimension of $\boldsymbol{\theta}$ and $\widehat{\theta}_j$ is the jth component of $\widehat{\boldsymbol{\theta}}$.

6. The CHS Study

The Cardiovascular Heath Study (CHS) is a population based observational prospective study of elderly adults (age ≥ 65 at enrollment) in the United States. A full description of the design of CHS is reported in Fried et al (1991). The analysis here includes 3967 subjects (1531 males) who were free of clinical cardiovascular events at enrollment.

The Framingham Risk Score (FR-score) is a widely used clinical prediction score used to quantify risk for cardiovascular events (Grundy et al, 1998, 2001). It includes information on age, cigarette smoking, blood pressure, diabetes mellitus, blood cholesterol and high density lipoproteins cholesterol. Separate score sheets are used for men and women. Here we consider the FR-score as a predictor for cardiovascular events, defined as non-fatal myocardial infarction or death due to coronary heart disease.

Subjects in this dataset were between 65 and 95 years old (mean 72.3 years) and free of clinical cardiovascular disease at enrollment. There were 585 (14.7%) who experienced a cardiovascular event. Follow up on subjects without cardiovascular events averaged 6.75 years (sd = 1.58 years). The FR-score was evaluated for all subjects at enrollment, and we only include this baseline assessment in this analysis. Although several other covariates were considered (including age and race), our analysis indicated that gender and medication for hypertension were covariates that had a substantial influence on the predictive accuracy of the FR-score. In particular we fit the

12

following models for the time dependent TPR and FPR functions:

$$\text{TPR}_{t,\mathbf{Z}}(y) = 1 - \Phi\{\alpha_1 t + \alpha_2 t^2 + \beta_1 Z_1 + \beta_2 Z_2 + h_0(y)\}$$

$$\text{FPR}_{t,\mathbf{Z}}(y) = 1 - \Phi\{b_1 Z_1 + b_2 Z_2 + c_0(y)\}$$

where $Z_1 = 1$ for subjects on medication for hypertension at enrollment and 0 otherwise, $Z_2 = 1$ for males and 0 otherwise, and we drop the subscript 's' relative to the general model forms (3) and (4) because Y, the FR-score, is measured only at baseline, $s = 0$. We choose $\tau = 7$ years in the analysis. That is, we investigated the predictive accuracy of the FR-score for events during the 7 years subsequent to enrollment. Subjects without an event by the $7th$ year of observation after the FR-score is measured are considered to be controls for the purposes of calculating the false positive rate.

There is little loss to follow up in CHS. However, 487 subjects in the sample died from other causes without a cardiovascular failure. Rather than censoring these survival times at death, we censor them at the end of the trial. This is appropriate since $1 - S_{\mathbf{Z}}(t)$ represents the marginal probability of a cardiovascular events by time t, which is not estimated by censoring at competing risk event times (Pepe and Mori, 1993).

The estimated regression coefficients and their estimated standard errors are shown in Table 2. The negative coefficient $\widehat{\beta}_1$ in the TPR model indicates that for a given positivity threshold y, medication use is associated with an increased TPR. That is, cases on medication had higher values of the FR-score than did cases not on medication. The coefficient $\widehat{\beta}_2 = 0.32$ indicates that male cases had lower values of the FR-score than did female cases. In contrast among the controls, gender had little effect on FR-score, $\widehat{b}_2 = 0.012$. Medication use among controls, like that in cases, was associated with higher FR-score values, $\widehat{b}_1 = -0.39$, the effect actually being greater in controls. Figure 3 displays the estimated TPR and FPR functions at $t = 1$ year and 5 years with their 95% confidence intervals and confidence bands for female subjects who are on hypertension medication. These curves indicate that for any positivity threshold y, the sensitivity of the FR-score is higher for events that occur at 1 year after enrollment than at 5 years after enrollment. In particular for these women, the threshold criterion FR-score > 10 which identifies 45% of subjects with events at 1 year, identifies only 36% at 5 years, while 30% of subjects who are event free by 7 years also meet this criterion. In Figure 4, we show ROC curves for different groups at $t = 1$

13

year and $t = 5$ years. Given the same false positive rate, the corresponding true positive rate is substantially higher in females than in males. Although for any given threshold, medication use is associated with increased TPR, it is not associated with increased accuracy of the FR-score. The ROC curves for subjects on medication are close to (a little below) those for subjects not on medication. This is because medication status affects the FPR functions to essentially the same degree as the TPR functions.

We next calculate the cumulative incidence based TPR and FPR functions. The curves for $\mathrm{TPR}_{t,\mathbf{z}}^{\mathsf{CI}}$ and $\mathrm{FPR}_{t,\mathbf{z}}^{\mathsf{CI}}$ are displayed in Figures 5 and 6. The definitions of the cumulative incidence based TPR and FPR functions specify cases and controls at time t as subjects with $T \leq t$ and subjects with $T > t$, respectively. At any given positivity threshold, both the true and false positive rates of the FR-score is also higher for events that occur within 1 year since enrollment than within 5 years since enrollment again indicating better discrimination for earlier events. In particular, for female subjects who are on hypertension medication, the threshold criterion FR-score which identifies 48% of subjects with events within 1 year, identifies only 39% of subjects with event within 5 years, while 32% of subjects who are event free by 1 year and 31% of subjects who are event free by 5 years meet this criterion.

7. Remarks

Statistical models for the joint analysis of longitudinal biomarkers and time to disease onset have been studied extensively in the past decade (e.g., Pawitan and Self, 1993; De Gruttola and Tu, 1994; Tsiatis, De Gruttola and Wulfsohn, 1995; Faucett and Thoams, 1996; Wulfsohn and Tsiatis 1997; Hogan and Laird, 1997; Hederson, Diggle and Dobson, 2000; Skates, Pauler and Jacobs, 2001; Wang and Taylor, 2001; Henderson, Diggle and Dobson, 2002). See Hogan and Laird (1997) for discussion of two broad classes of models, namely selection models and pattern mixture models. Most of the existing methods in this area require parametric modelling of the marker process over time and a joint parametric model for the distribution of the event time. To induce models for the association between the marker and event time process, both mixture models and selection models rely on specication of the distributional assumptions for random effects or latent stochastic processes. In contrast, we use marginal semi-parametric models for the marker distribution given the event time and for the event time distribution. The approach does not model marker processes

14

and hence is more flexible. We estimate the regression parametere and the non-parametric baseline functions simultaneously based on estimating equations and incorporate censoring by integrating over time.

In this article, we are only concerned with the distribution of biomarkers measured prior to diagnosis, the biomarker measured after $\min(T, \tau)$ does not contribute to estimating the parameters in model (3) and (4). We delete these observations from the data for analysis. However, the methods could be extended to include biomarkers measured after T. In some settings, one may be interested in using the biomarker to detect T as soon as possible after the event occurs.

The covariate effects are modeled separately for the TPR and FPR functions. Therefore, the regression parameter, β in TPR or b in FPR, do not directly represent covariate effects on the ROC curve of the biomarker for detecting disease. It would be interesting to investigate if one could extend the regression models for time dependent ROC curves such as those proposed by Etzioni et al (1999) and Cai and Pepe (2002) in order to incorporate censored data into the analysis.

Acknowledgement

15

Appendix

A   Uniqueness and Consistency

We first show that $\widetilde{\psi}$ is a consistent estimate of ψ_0 for $y \in [\mathfrak{l}, \mathfrak{u}]$. To this end, let

$$\ell(y; \boldsymbol{\psi}) = \sum_{\Delta_i = 0} \sum_{k=1}^{K_i} \left[I(Y_{ik} \geq y) \log p_{ik}(y; \boldsymbol{\psi}) + \{1 - I(Y_{ik} \geq y)\} \log \{1 - p_{ik}(y; \boldsymbol{\psi})\} \right]$$

and let $[\widetilde{h}(y; \boldsymbol{\theta}), \widetilde{c}(y; \boldsymbol{\theta})]'$ denote the solution to (8) at a given $\boldsymbol{\theta}$. It is easy to see that

$$\begin{bmatrix} \widetilde{h}(y; \boldsymbol{\theta}) \\ \widetilde{c}(y; \boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \operatorname{argmax}_{h(y)} \ell(y; \boldsymbol{\psi}) \\ \operatorname{argmax}_{c(y)} \ell(y; \boldsymbol{\psi}) \end{bmatrix}, \qquad \text{and} \qquad \widetilde{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \widetilde{Q}(\boldsymbol{\theta}),$$

where $\widetilde{Q}(\boldsymbol{\theta}) = \int_{\mathfrak{l}}^{\mathfrak{u}} \ell \left\{ y; \widetilde{h}(\cdot; \boldsymbol{\theta}), \widetilde{c}(\cdot; \boldsymbol{\theta}), \boldsymbol{\theta} \right\} d\widehat{v}(y)$. Suppose $\boldsymbol{\theta}_0$ lies in a compact set $\Re_\theta$. To show that $\widetilde{\boldsymbol{\theta}}$, the maximizer of $\widetilde{Q}(\boldsymbol{\theta})$ is strongly consistent, it is sufficient to show that $\widetilde{Q}(\boldsymbol{\theta})$ converges uniformly to a deterministic function of $\boldsymbol{\theta}$ almost surely, which has a unique minimizer at $\boldsymbol{\theta}_0$ (Newey and McFadden, 1994). To this end, let $\mathbb{L}(\cdot)$ denote the distribution function of $\mathcal{F} = \{I(X \leq \tau), \delta, \min(X, \tau), s, \mathbf{Z}\}$, $\Re_{\widehat{v}} = \{y \in [\mathfrak{l}, \mathfrak{u}] : d\widehat{v}(y)/dy \neq 0\}$, $\widetilde{h}(y) = \widetilde{h}(y; \widetilde{\boldsymbol{\theta}})$ and $\widetilde{c}(y) = \widetilde{c}(y; \widetilde{\boldsymbol{\theta}})$. Note that,

$$\begin{bmatrix} \left\{ \widetilde{h}(y), \widetilde{c}(y) \right\} : y \in \Re_{\widehat{v}}, \\ \widetilde{\boldsymbol{\theta}} \end{bmatrix} = \operatorname*{argmax}_{\{\boldsymbol{\psi} : y \in \Re_{\widehat{v}}\}} \int \ell(y; \boldsymbol{\psi}) d\widehat{v}(y).$$

It follows from strong law of large numbers that $n^{-1} \int_{\mathfrak{l}}^{\mathfrak{u}} \ell(y; \boldsymbol{\psi}) d\widehat{v}(y) \to \int_{\mathfrak{l}}^{\mathfrak{u}} q(y; \boldsymbol{\psi}) dv(y)$, where

$$q(y; \boldsymbol{\psi}) = \mathcal{K}_0 \int p_{11}(y; \boldsymbol{\psi}_0) \log p_{11}(y; \boldsymbol{\psi}) + \{1 - p_{11}(y; \boldsymbol{\psi}_0)\} \log \{1 - p_{11}(y; \boldsymbol{\psi})\} d\mathbb{L}(\mathcal{F}_{11}),$$

and $\mathcal{K}_0 = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n K_i(1 - \Delta_i)$. Since $f(p) = p_0 \log(p) + (1 - p_0) \log(1 - p)$ has a unique maximizer $p = p_0$ for any $p_0 \in (0, 1)$, we have $q(y; \boldsymbol{\psi}) \geq q(y; \boldsymbol{\psi}_0)$ and the equality holds if and only if $p_{11}(y; \boldsymbol{\psi}) = p_{11}(y; \boldsymbol{\psi}_0)$, for $y \in \Re_v = \{y \in [\mathfrak{l}, \mathfrak{u}] : dv(y)/dy > 0\}$ and $\mathcal{F}_{11} \in \Re_{\mathcal{F}}$ where $\Re_{\mathcal{F}}$ is the domain of $\mathcal{F}_{11}$. It follows from the monotonicity of $g(\cdot)$ and $g_\tau(\cdot)$ that

$$h(y) + \boldsymbol{\beta}'\mathbf{z} + \eta_{\boldsymbol{\alpha}}(x, s) = h_0(y) + \boldsymbol{\beta}_0'\mathbf{z} + \eta_{\boldsymbol{\alpha}_0}(x, s) \tag{A.1}$$

$$c(y) + \mathbf{b}'\mathbf{z} + \xi_{\boldsymbol{a}}(s) = c_0(y) + \mathbf{b}_0'\mathbf{z} + \xi_{\boldsymbol{a}_0}(s) \tag{A.2}$$

for $y \in \Re_v$ and any $(x, s, \mathbf{z})$ in the subspace of $\Re_{\mathcal{F}}$ restricting to $\Delta = 0$. Without loss of generality, we assume that $\Re_{\mathcal{F}}$ is not degenerate. It follows that (A.1) and (A.2) holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

16

$h(y) = h_0(y)$, and $c(y) = c_0(y)$ for $y \in \Re_v$. This ensures that the limit of $\widetilde{Q}(\boldsymbol{\theta})$, denoted by $q(\boldsymbol{\theta})$, has a unique minimizer at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. It is not hard to show that that $\widetilde{Q}(\boldsymbol{\theta})$ is equicontinuous in $\boldsymbol{\theta}$, therefore the convergence of $\widetilde{Q}(\boldsymbol{\theta}) \to q(\boldsymbol{\theta})$ is uniform in $\boldsymbol{\theta} \in \Re_\theta$. This concludes the arguments for consistency of $\widetilde{\boldsymbol{\theta}}$. The equicontinuity of $\ell(y; \boldsymbol{\theta}_0)$ in $h(y)$ and $c(y)$ ensures the consistency of $\widetilde{h}(y)$ and $\widetilde{c}(y)$. The consistency of $\widetilde{h}(\cdot)$ and $\widetilde{c}(\cdot)$ is uniform in $y \in [\mathfrak{l}, \mathfrak{u}]$ since $h(\cdot)$ and $c(\cdot)$ are monotone functions of y.

It remains to show the consistency of $\widehat{\boldsymbol{\psi}}$. Let $\mathbf{V}^\mathbf{H}(y; \boldsymbol{\psi})$ and $\mathbf{V}(\boldsymbol{\theta})$ denote the left hand side of (8) and (9), respectively. It follows from the strong law of large numbers and the uniform consistency of $\widehat{S}_{\mathbf{z}}(\cdot)$ that for any $\mu \geq 0$, $y \in [\mathfrak{l}, \mathfrak{u}]$, $\boldsymbol{\theta} \in \Re_\mu = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \mu\}$, and sufficiently large ϵ,

$$\mathbf{V}^\mathbf{H}(y; \boldsymbol{\psi}_{+\epsilon}) > 0, \quad \mathbf{V}^\mathbf{H}(y; \boldsymbol{\psi}_{-\epsilon}) < 0 \tag{A.3}$$

where $\boldsymbol{\psi}_{\pm\epsilon} = \{\mathbf{H}(\cdot) \pm \epsilon; \boldsymbol{\theta}\}$. This, coupled with the monotonicity and continuity of g and g_τ, implies that there exists a unique $\widehat{\mathbf{H}}(\cdot; \boldsymbol{\theta})$ such that

$$\mathbf{V}^\mathbf{H}\left\{y; \widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}}\right\} = 0, \tag{A.4}$$

where $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}} = \left\{\widehat{\mathbf{H}}(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta}\right\}$. Since (A.3) holds for any $\epsilon > 0$ only for the case that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\widehat{\mathbf{H}}(y; \boldsymbol{\theta}_0) \to \mathbf{H}_0(y)$, uniformly in $y \in [\mathfrak{l}, \mathfrak{u}]$. To show the consistency of $\widehat{\boldsymbol{\theta}}$, we first note that $n^{-1}\partial\mathbf{V}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = \widehat{\mathbb{A}}(\boldsymbol{\theta})$, where

$$\widehat{\mathbb{A}}(\boldsymbol{\theta}) = -\sum_{i=1}^n \sum_{k=1}^{K_i} \int_\mathfrak{l}^\mathfrak{u} \widehat{\mathbf{W}}_{ik}^{\boldsymbol{\theta}}(y; \widetilde{\boldsymbol{\psi}}) \left\{\widehat{\boldsymbol{\omega}}_{ik}^\mathbf{H}(y; \widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}})' \frac{\partial\widehat{\mathbf{H}}(y; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \widehat{\boldsymbol{\omega}}_{ik}^{\boldsymbol{\theta}}(y; \widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}})'\right\},$$

$\widehat{\boldsymbol{\omega}}_{ik}^\mathbf{H}(y; \boldsymbol{\psi}) = \frac{\partial\widehat{p}_{ik}}{\partial\mathbf{H}}(y; \boldsymbol{\psi})$, and $\widehat{\boldsymbol{\omega}}_{ik}^{\boldsymbol{\theta}}(y; \boldsymbol{\psi}) = \frac{\partial\widehat{p}_{ik}}{\partial\boldsymbol{\theta}}(y; \boldsymbol{\psi})$. By differentiating both sides of (A.4) with respect to $\boldsymbol{\theta}$, we obtain the identity

$$\frac{\partial\widehat{\mathbf{H}}(y; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = -\left\{\sum_{i=1}^n \sum_{k=1}^{K_i} \widehat{\mathbf{W}}_{ik}^\mathbf{H}(y; \widetilde{\boldsymbol{\psi}})\widehat{\boldsymbol{\omega}}_{ik}^\mathbf{H}(y; \widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}})'\right\}^{-1} \left\{\sum_{i=1}^n \sum_{k=1}^{K_i} \widehat{\mathbf{W}}_{ik}^\mathbf{H}(y; \widetilde{\boldsymbol{\psi}})\widehat{\boldsymbol{\omega}}_{ik}^{\boldsymbol{\theta}}(y; \widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}})\right\}. \tag{A.5}$$

This, coupled with the consistency of $\widetilde{\boldsymbol{\psi}}$, $\widehat{\mathbf{H}}(\cdot; \boldsymbol{\theta}_0)$, implies that $n^{-1}\partial\mathbf{V}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}' \to \mathbb{A}$, where

$$\mathbb{A} = -\int_\mathfrak{l}^\mathfrak{u} \left[\mathbb{W}^{\boldsymbol{\theta\theta}}(y) - \mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y)' \{\mathbb{W}^{\mathbf{HH}}(y)\}^{-1} \mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y)\right] dv(y), \tag{A.6}$$

$$\mathbb{W}^{\boldsymbol{\theta\theta}}(y) = E\left\{\mathbb{W}_{ik}^{[02]}(y; \boldsymbol{\psi}_0)\right\}, \qquad \mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y) = E\left\{\mathbb{W}_{ik}^{[11]}(y; \boldsymbol{\psi}_0)\right\},$$

$$\mathbb{W}^{\mathbf{HH}}(y) = E\left\{\mathbb{W}_{ik}^{[20]}(y; \boldsymbol{\psi}_0)\right\}, \qquad \mathbb{W}_{ik}^{[ml]}(y; \boldsymbol{\psi}) = \frac{\frac{\partial p_{ik}}{\partial\mathbf{H}}(y; \boldsymbol{\psi})^{\otimes m} \left\{\frac{\partial p_{ik}}{\partial\boldsymbol{\theta}}(y; \boldsymbol{\psi})^{\otimes l}\right\}'}{p_{ik}(y; \boldsymbol{\psi})\{1 - p_{ik}(y; \boldsymbol{\psi})\}},$$

17

and we use the notations that for a vector $\mathbf{x}$, $\mathbf{x}^{\otimes 0} = 1$, $\mathbf{x}^{\otimes 1} = \mathbf{x}$ and $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}'$. When $\vec{\mathbf{Z}}_{ik}$ and $\mathbf{Z}_{ik}$ are non-degenerate, $\mathbb{A}$ is positive definite. Now, since $n^{-1}\mathbf{V}(\boldsymbol{\theta}_0) \to 0$, by the standard inverse function theorem, there exists a unique solution $\widehat{\boldsymbol{\theta}}$ to the equation $\mathbf{V}(\boldsymbol{\theta}) = 0$ in a neighborhood of $\boldsymbol{\theta}_0$. This also implies that $\widehat{\boldsymbol{\theta}}$ is strongly consistent and $\widehat{\mathbf{H}}(y;\widehat{\boldsymbol{\theta}}) \to \mathbf{H}_0(y)$, almost surely, uniformly in $y \in [\mathfrak{l},\ \mathfrak{u}]$.

B   Large sample distribution of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{H}}(y)$

By the consistency of $\widehat{\boldsymbol{\theta}}$ and a Taylor's series expansion of $\mathbf{V}(\widehat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$, we obtain

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \approx \mathbb{A}^{-1} n^{\frac{1}{2}} \mathbf{V}(\boldsymbol{\theta}_0).$$

It follows from a Taylor's series expansion of $\widehat{\mathbf{H}}(y;\boldsymbol{\theta}_0)$ around $\mathbf{H}_0(y)$ and the uniform consistency of $\widehat{S}_{\mathbf{z}}(\cdot)$ that

$$n^{-\frac{1}{2}}\mathbf{V}(\boldsymbol{\theta}_0) \approx n^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{k=1}^{K_i} \int_{\mathfrak{l}}^{\mathfrak{u}} \mathbf{v}_{ik}^{\boldsymbol{\theta}}(y) \left[e_{ik}(y) + \int_{X_i}^{\tau} g\left\{\boldsymbol{\theta}_{10}'\vec{\mathbf{Z}}_{ik}^{[u]} + h_0(y)\right\} d\left\{ \frac{S_{\mathbf{Z}_i}(u)}{S_{\mathbf{Z}_i}(X_i)} - \frac{\widehat{S}_{\mathbf{Z}_i}(u)}{\widehat{S}_{\mathbf{Z}_i}(X_i)} \right\} \right.$$
$$\left. + g_\tau\left\{\boldsymbol{\theta}_{20}'\mathbf{Z}_{ik} + c_0(y)\right\} \left\{ \frac{S_{\mathbf{Z}_i}(\tau)}{S_{\mathbf{Z}_i}(X_i)} - \frac{\widehat{S}_{\mathbf{Z}_i}(\tau)}{\widehat{S}_{\mathbf{Z}_i}(X_i)} \right\} \right]$$
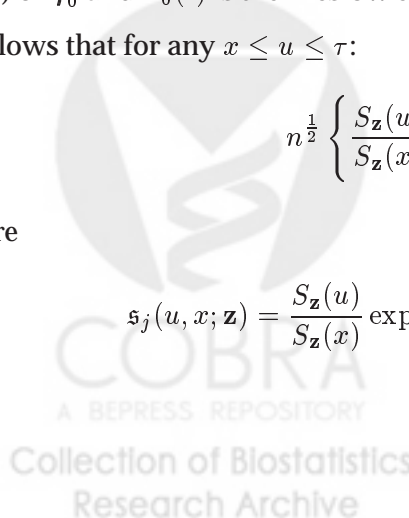
where

$$\mathbf{v}_{ik}^{\boldsymbol{\theta}}(y) = \mathbf{W}_{ik}^{\boldsymbol{\theta}}(y;\boldsymbol{\psi}_0) - \mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y)' \left\{\mathbb{W}^{\mathbf{H}\mathbf{H}}(y)\right\}^{-1} \mathbf{W}_{ik}^{\mathbf{H}}(y;\boldsymbol{\psi}_0). \tag{B.7}$$

Recall that $\widehat{S}_{\mathbf{z}}(t) = \exp\{-\widehat{\Lambda}_0(t)\exp(\widehat{\boldsymbol{\gamma}}'\mathbf{z})\}$, where $\widehat{\boldsymbol{\gamma}}$ is the MPLE (maximum partial likelihood estimate) of $\boldsymbol{\gamma}_0$ and $\widehat{\Lambda}_0(t)$ is the Breslow estimate for $\Lambda_0(t)$. From standard theory for survival analysis, it follows that for any $x \le u \le \tau$:

$$n^{\frac{1}{2}} \left\{ \frac{S_{\mathbf{z}}(u)}{S_{\mathbf{z}}(x)} - \frac{\widehat{S}_{\mathbf{z}}(u)}{\widehat{S}_{\mathbf{z}}(x)} \right\} \approx n^{-\frac{1}{2}} \sum_{j=1}^{n} \mathfrak{s}_j(u,x;\mathbf{z}),$$

where

$$\mathfrak{s}_j(u,x;\mathbf{z}) = \frac{S_{\mathbf{z}}(u)}{S_{\mathbf{z}}(x)} \exp(\boldsymbol{\gamma}_0'\mathbf{z}) \left[\{\mathbf{O}(u;\mathbf{z}) - \mathbf{O}(x;\mathbf{z})\}' \boldsymbol{\ell}_i + \int_x^u \frac{dM_i(s)}{\mathbb{S}^{(0)}(s)} \right], \tag{B.8}$$

18

$$\mathbf{O}(u;\mathbf{z}) = \lim_{n\to\infty} \left\{ \int_0^\infty \frac{\mathcal{S}^{(2)}(t)\mathcal{S}^{(0)}(t) - \mathcal{S}^{(1)}(t)^{\otimes 2}}{\mathcal{S}^{(0)}(t)^2} d\bar{N}(t) \right\}^{-1} \left\{ \Lambda_0(u)\mathbf{z} + \int_0^u \frac{\mathcal{S}^{(1)}(s)}{\mathcal{S}^{(0)}(s)^2} d\bar{N}(s) \right\},$$

$$\ell_i = \int_0^\infty \left\{ \mathbf{Z}_i - \frac{\mathcal{S}^{(1)}(t)}{\mathcal{S}^{(0)}(t)} \right\} dM_i(t), \qquad\qquad M_i(t) = N_i(t) - \int_0^t I(X_i \geq s) e^{\boldsymbol{\gamma}_0' \mathbf{Z}_i} d\Lambda_0(s),$$

$$\bar{N}(t) = \sum_{i=1}^n N_i(t)/n = \sum_{i=1}^n I(X_i \leq t)\delta_i/n, \qquad \mathcal{S}^{(k)}(t) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n I(X_i \geq t) \exp(\boldsymbol{\gamma}_0' \mathbf{Z}_i) \mathbf{Z}_i^{\otimes k}.$$

As a consequence,

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \approx n^{-\frac{3}{2}} \mathbb{A}^{-1} \sum_{i=n_d+1}^{n_\tau} \sum_{j=1}^n \mathbf{U}_{ij}$$

To derive the large sample distribution of $n^{\frac{1}{2}} \left\{ \widehat{\mathbf{H}}(y) - \mathbf{H}_0(y) \right\}$, we take a Taylor's series expansion of $\widehat{\mathbf{H}}(y) = \widehat{\mathbf{H}}(y; \widehat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$ and obtain

$$n^{\frac{1}{2}} \left\{ \widehat{\mathbf{H}}(y) - \mathbf{H}_0(y) \right\} \approx n^{-\frac{3}{2}} \sum_{i=n_d+1}^{n_\tau} \sum_{j=1}^n \mathcal{H}_{ij}(y)$$

where

$$\mathcal{H}_{ij}(y) = \begin{bmatrix} \mathfrak{h}_{ij}(y) \\ \mathfrak{c}_{ij}(y) \end{bmatrix} = \mathbb{W}^{\mathbf{H}\mathbf{H}}(y)^{-1} \left\{ \mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y) \mathbf{U}_{ij} + \mathcal{E}_{ij}(y; \mathbf{W}^{\mathbf{H}}) + \sum_{l=1}^{K_j} \mathbf{W}_{jl}^{\mathbf{H}}(y; \boldsymbol{\psi}_0) e_{jl}(y) \right\}. \qquad \text{(B.9)}$$

It follows that

$$\mathbf{Q}(y; t, s, \mathbf{z}) \approx n^{-\frac{3}{2}} \sum_{i=n_d+1}^{n_\tau} \sum_{j=1}^n \left\{ \mathcal{H}_{ij}(y) + \begin{bmatrix} \boldsymbol{\eta}(t,s)' & \mathbf{z}' & 0 & 0 \\ 0 & 0 & \boldsymbol{\xi}(s)' & \mathbf{z}' \end{bmatrix} \mathbb{A}^{-1} \mathbf{U}_{ij} \right\}$$

REFERENCES

BALASUBRAMANIAN, R. & LAGAKOS, S. (2001). Estimation of the timing of perinatal transmission of hiv. *Biometrics* **57**, 1048–58.

CAI, T. & PEPE, M. S. (2002). Semi-parametric roc analysis to evaluate biomarkers for disease. *J. Am. Statist. Assoc.* in press.

CAI, T., WEI, L. J. & WILCOX, M. (2000). Semi-parametric regression analysis for clustered failure time data. *Biometrika* **87**. 867–78.

CHENG, S. C., WEI, L. J. & YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–45.

CHENG, S. C., WEI, L. J. & YING, Z. (1997). Prediction of survival probabilities with semi-parametric transformation models. *J. Am. Statist. Assoc.* **92**, 227–35.

DABROWSKA, D. & DOKSUM, K. (1988a). Estimation and testing in a two-sample generalized odds-rate model. *J. Am. Statist. Assoc.* **83**, 744–9.

DABROWSKA, D. & DOKSUM, K. (1988b). Partial likelihood in transformation models with censored data. *Scand. J. Statist.* **15**, 1–23.

ETZIONI, R., PEPE, M. S., LONGTON, G., HU, C. & GOODMAN, G. (1999). Incorporating the time dimension in Receiver Operating Characteristic curves: A case study of prostate cancer. *Medical Decision Making* **19**, 242–251.

FAUCETT, C. L. & THOMAS, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley.

FRIED, L. P., BORHANI, N., ENRIGHT, P., FURBERG, C., GARDIN, J., KRONMAL, R., KULLER, L., MANOLIO, T., MITTELMARK, M., NEWMAN, A., DH, O., PSATY, B., RAUTAHARJU, P., TRACY, R. & WEILER, P. (1991). The cardiovascular health study: design and rationale. *Ann Epidemiol* **1**, 263–76.

GRUNDY, S., BALADY, G., CRIQUI, M., FLETCHER, G., GREENLAND, P., HIRATZKA, L., HOUSTON-MILLER, N., KRIS-ETHERTON, P., KRUMHOLZ, H., LAROSA, J., OCKENE, I., PEAR-

SON, T., REED, J., WASHINGTON, R. & SMITH, S. (1998). Primary prevention of coronary heart disease: Guidance from framingham - a statement for healthcare professionals from the aha task force on risk reduction. *Circulation* **97**, 1876–87.

GRUNDY, S., D'AGOSTINO, R., MOSCA, L., BURKE, G., WILSON, P., RADER, D., CLEEMAN, J., ROCCELLA, E., CUTLER, J. & FRIEDMAN, L. (2001). Cardiovascular risk assessment based on us cohort studies - findings from a national heart, lung, and blood institute workshop. *Circulation* **104**, 491–96.

HEAGERTY, P. J., LUMLEY, T. & S., P. M. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337–44.

HENDERSON, R., DIGGLE, P. & DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–80.

HENDERSON, R., DIGGLE, P. & DOBSON, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics* **3**, 33–50.

HOGAN, J. W. & LAIRD, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259–272.

LEISENRING, W., PEPE, M. S. & LONGTON, G. L. (1997). A marginal regression modelling framework for evaluating medical diagnostic tests. *Statistics in Medicine* **16**, 1263–81.

NEWEY, W. & MCFADDEN, D. (1994). Estimation in large samples. In McFadden, D. & Engler, R., editors, *In Handbook of Econometric*, volume 4. Amsterdam: North Holland.

PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

PAWITAN, Y. & SELF, S. (1993). Modeling disease marker processes in Aids. *J. Am. Statist. Assoc.* **88**, 719–726.

PEPE, M. S. (2003). *Statistical Evaluation of Diagnostic Tests and Biomarkers*. Oxford University Press.

PEPE, M. S., ETZIONI, R., FENG, Z. D., POTTER, J., THOMPSON, M., THORNQUIST, M., WINGET, M. & YASUI, Y. (2001). Phases of biomarker development for early detection of cancer. *J. Natl. Cancer I* **93**, 1054–61.

PEPE, M. S. & MORI, M. (1993). Kaplan-meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* **12**, 737–751.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* Hayward, CA: Institute of Mathematical Statistics.

SCHARFSTEIN, D. O., TSIATIS, A. A. & GILBERT, P. (1998). Semi-parametric efficient estimation in the generalized odds-rate class of regression models for right-censored time to event data. *Lifetime Data Anal.* **4**, 355–91.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

SKATES, S. J., PAULER, D. K. & JACOBS, I. J. (2001). Screening based on the risk of cancer calculation from bayesian hierarchical change-point models of longitudinal markers. *J. Am. Statist. Assoc.* **96**, 429–39.

TSIATIS, A. A., DEGRUTTOLA, V. & WULFSOHN, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and Cd4 counts in patients with Aids. *J. Am. Statist. Assoc.* **90**, 27–37.

VAN'T VEER, L., DAI, H. & ET AL (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–35.

WANG, Y. & TAYLOR, J. M. G. (2001). Joint modelling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Am. Statist. Assoc.* **96**, 895–905.

WILSON, P., D'AGOSTINO, R., LEVY, D., BELANGER, A., SILBERSHATZ, H. & KANNEL, W. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–47.

WULFSOHN, M. S. & TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

Table 1: Bias and mean squared error (MSE) of $\widehat{\alpha}$ and $\widehat{\beta}$ at sample size $n = 200$ with $K = 1$ and $K = 2$. Results are shown for 1000 simulated data sets.

(a) Results for the first setting where about 50% of subjects have observations times $X > \tau$, and 28% of subjects are censored before τ.

| | $K = 1$ | | | | | | $K = 2$ | | | | | |
| | Bias | | | MSE | | | Bias | | | MSE | | |
ρ	$\alpha = .5$	$\beta = 1$	$b = 1$	α	β	b	$\alpha = .5$	$\beta = 1$	$b = 1$	α	β	b
0	.040	.085	.010	.026	.158	.039	.021	.036	.011	.014	.073	.024
.5	.040	.087	.013	.026	.156	.033	.021	.036	.010	.014	.071	.020
1	.040	.089	.016	.026	.154	.031	.021	.036	.010	.014	.071	.019
2	.041	.094	.022	.026	.154	.032	.022	.037	.011	.014	.070	.020
5	.044	.107	.036	.027	.164	.046	.024	.041	.014	.015	.075	.026

(b) Results for the second setting where about 37% of subjects have observation times $X > \tau$, 50% of subjects are censored before τ.

| | $K = 1$ | | | | | | $K = 2$ | | | | | |
| | Bias | | | MSE | | | Bias | | | MSE | | |
ρ	$\alpha = .5$	$\beta = 1$	$b = 1$	α	β	b	$\alpha = .5$	$\beta = 1$	$b = 1$	α	β	b
0	.060	.119	.023	.060	.267	.049	.033	.063	.015	.020	.121	.033
.5	.063	.118	.024	.060	.248	.038	.033	.065	.017	.020	.111	.024
1	.064	.119	.027	.060	.239	.036	.034	.068	.019	.020	.107	.022
2	.068	.122	.032	.061	.234	.038	.036	.072	.021	.020	.104	.022
5	.080	.136	.040	.067	.258	.050	.043	.086	.023	.022	.118	.027

23

Table 2: Estimated regression coefficients and their estimated standard errors for the TPR and FPR models of the Framingham Risk Score as a predictor for cardiovascular events within 7 years.

		α_1	α_2	β_1	β_2	b_1	b_2
$\rho = 1$	Estimate	0.190	-0.021	- 0.275	0.318	-0.390	0.012
	Std. Error	0.078	0.011	0.086	0.087	0.037	0.033
$\rho = 0$	Estimate	0.190	-0.021	-0.274	0.317	-0.388	0.019
	Std. Error	0.078	0.011	0.086	0.087	0.039	0.035

Figure 1: Sample averages (thiner curves) of the estimated true and false positive rate functions for $Y(0)$ at time $t = 2$ as functions of the threshold y for positivity compared to the truth (thicker curves). The covariate value z is chosen as 1. Solid curves are for incidence based TPR and FPR and dashed curves are for cumulative incidence based TPR and FPR functions. The results are based on 1000 simulated datasets with sample size $n = 200$, $K = 2$, and the weight ρ is set at 1.

(a) TPR

(b) FPR

25
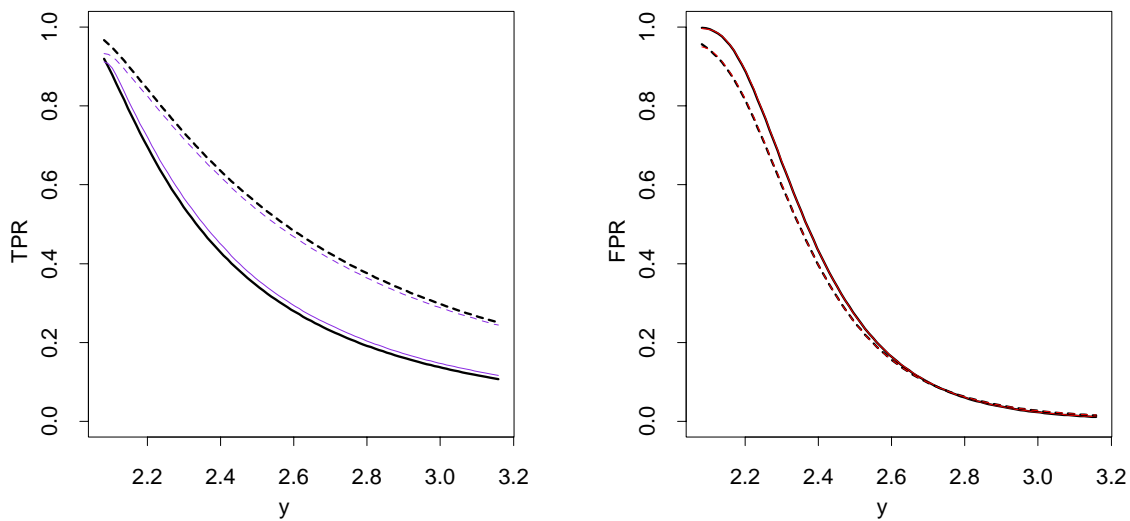
Figure 2: Sample averages (thiner curves) of the estimated true and false positive functions for $Y(0)$ using the threshold $y = 2$ as a function of event time t, compared to the truth (thicker curves). The covariate z is chosen to be 1. Solid curves are for incidence based TPR and FPR and dashed curves are for cumulative incidence based TPR and FPR functions. The results are based on 1000 simulated datasets with sample size $n = 200$, $K = 2$, and the weight ρ is set at 1.

(a) TPR

(b) FPR

26
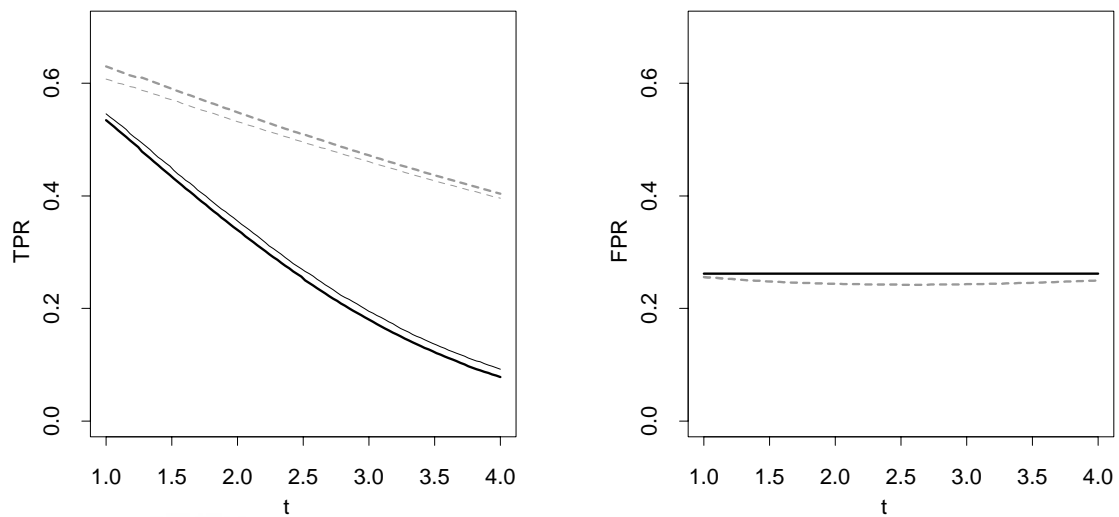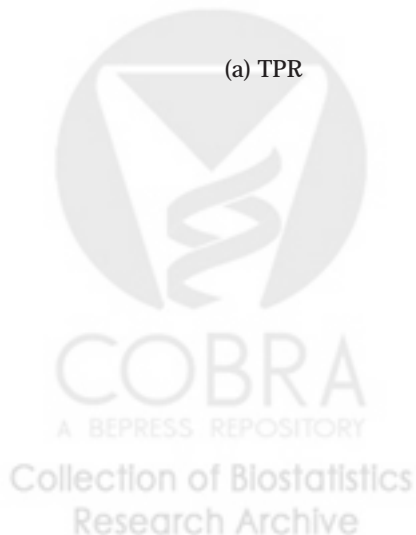
Figure 3: Estimated TPR functions of the Framingham risk score for female subjects who are on medication. The sensitivities (TPR) for events at $t = 1$ and $t = 5$ years after the FR-score is measured are displayed. Shown also are their 95% simultaneous confidence bands.
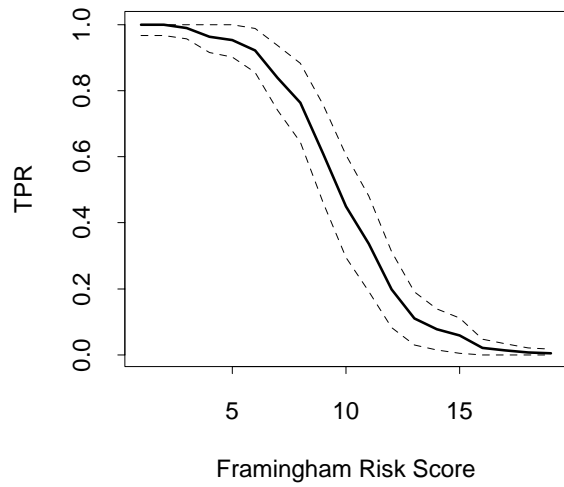
(a) TPR : $T = 1$ year
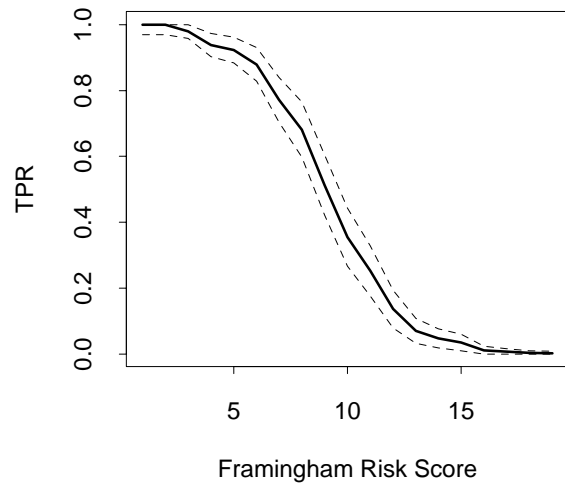
(b) TPR : $T = 5$ year
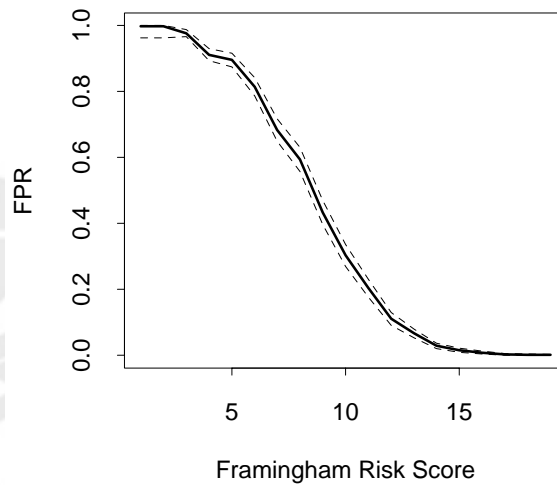
(c) FPR : $T > \tau = 7$ years

Figure 4: Plots of the ROC curve: the TPR function vs the FPR function at $t = 1$ year and $t = 5$ years. Shown are plots for females (solid curves) and for males (dashed curves). Thicker curves are for those on medication and thiner curves are for those not on medication.
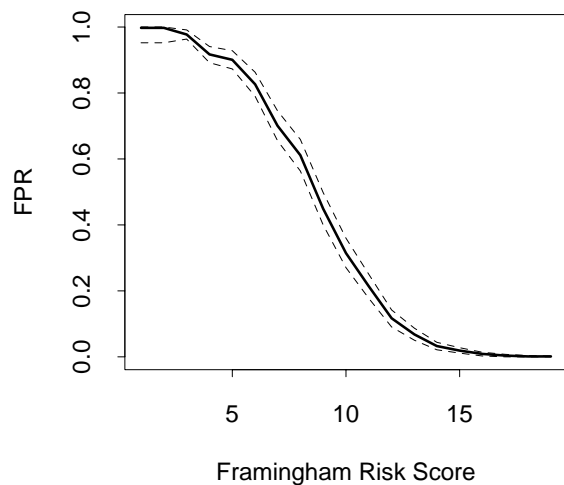

(a) $t = 1$ year

(b) $t = 5$ years

28
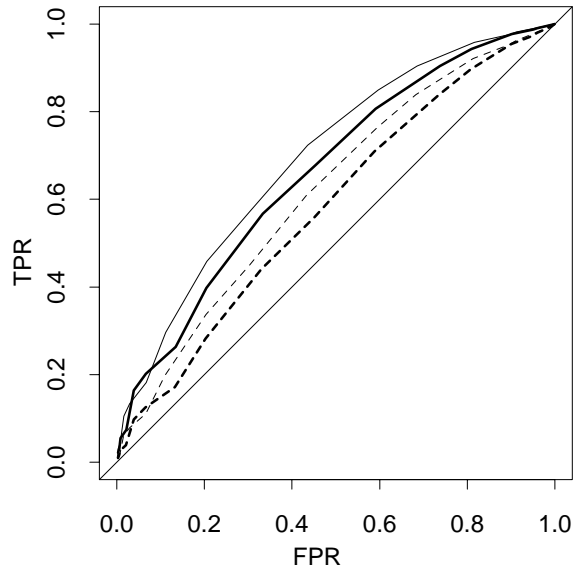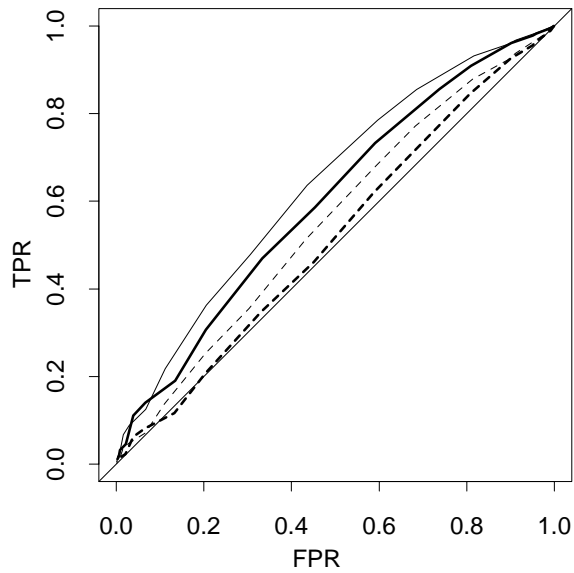
Figure 5: Estimated cumulative incidence based TPR and FPR functions of the Framingham risk score for female subjects who are on medication. Shown also are their 95% simultaneous confidence bands.
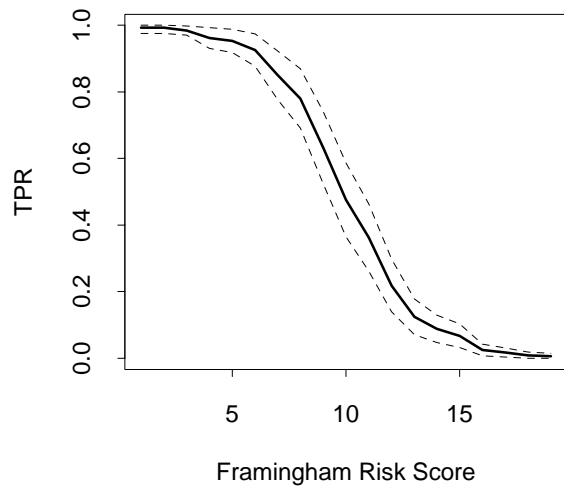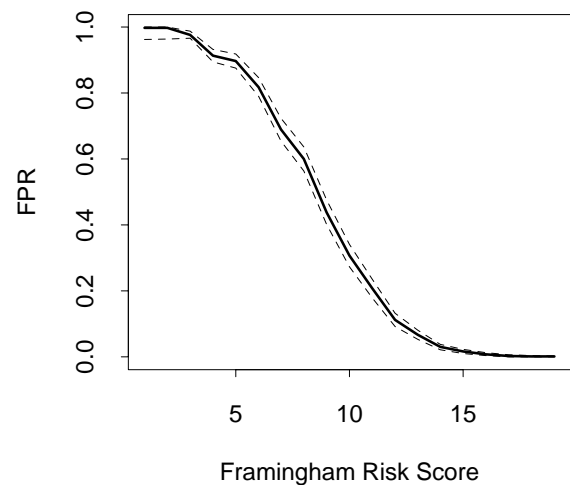

(a) TPR$^{\text{CI}}$: $T \leq 1$ year
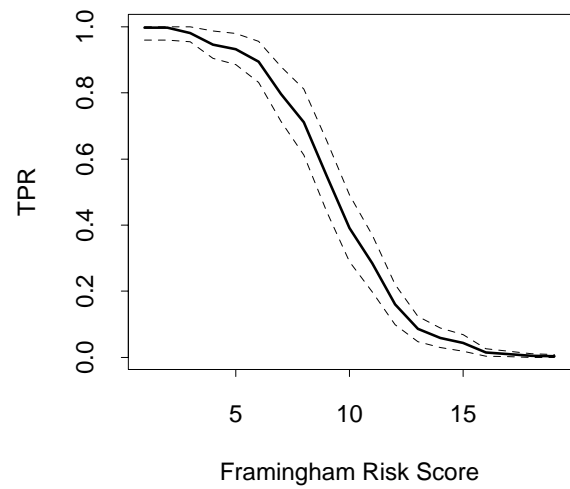
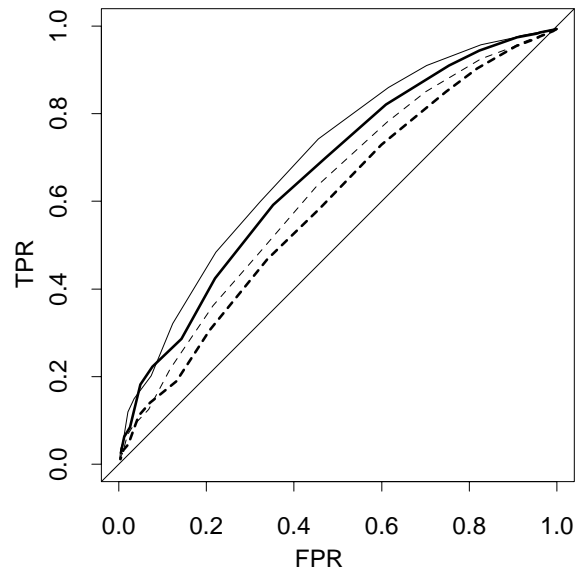(b) TPR$^{\text{CI}}$: $T \leq 5$ years

(c) FPR$^{\text{CI}}$: $T > 1$ year

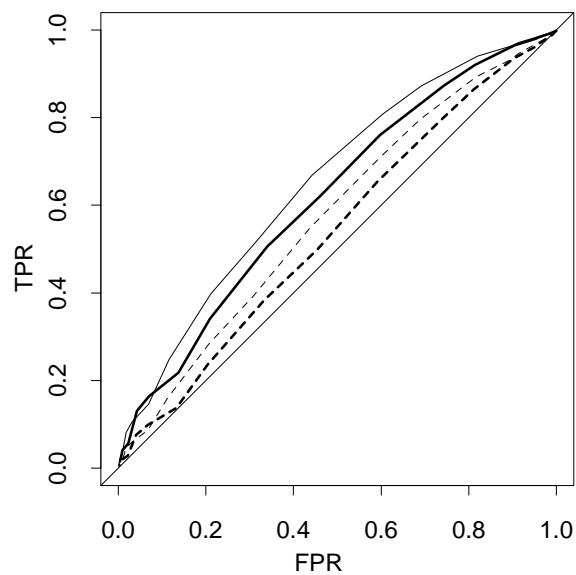(d) FPR$^{\text{CI}}$: $T > 5$ years

Figure 6: Plots of the cumulative incidence based ROC curve at $t = 1$ year and $t = 5$ years. Shown are plots for females (solid curves) and for males (dashed curves). Thicker curves are for those on medication and thiner curves are for those not on medication.

(a) $t = 1$ year

(b) $t = 5$ years